



**HAL**  
open science

# Ideophones: honing in on a descriptive and typological concept

Aimée Lahaussais, Julie Marsault, Yvonne Treis

► **To cite this version:**

Aimée Lahaussais, Julie Marsault, Yvonne Treis. Ideophones: honing in on a descriptive and typological concept. *Linguistic Typology at the Crossroads*, 4 (1), 2024. halshs-04612053

**HAL Id: halshs-04612053**

<https://shs.hal.science/halshs-04612053v1>

Submitted on 17 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Linguistic Typology

---

# at the Crossroads



ISSN 2785-0943

Volume 4 – Issue 1 – 2024

Issue DOI: <https://doi.org/10.6092/issn.2785-0943/v4-n1-2024>

This journal provides immediate and free open access. There is no embargo on the journal's publications. Submission and acceptance dates, along with publication dates, are made available on the PDF format for each paper. The authors of published articles remain the copyright holders and grant third parties the right to use, reproduce, and share the article according to the [Creative Commons Attribution 4.0 International](#) license agreement. The reviewing process is double-blind. Ethical policies and indexing information are publicly available on the journal website:

<https://typologyatcrossroads.unibo.it>

## Editors

*Nicola Grandi* (University of Bologna, Editor in chief)

*Caterina Mauri* (University of Bologna, Editor in chief)

*Francesca Di Garbo* (University of Aix-Marseille)

*Andrea Sansò* (University of Insubria)

## Publisher

*Department of Classical Philology and Italian Studies* (University of Bologna)

*Department of Modern Languages, Literatures and Cultures* (University of Bologna)

The journal is hosted and maintained by [AlmaDL](#)



# Linguistic Typology

---

## at the Crossroads



### Editorial board

*Mira Ariel* (Tel Aviv University)  
*Sonia Cristofaro* (Sorbonne Université)  
*Chiara Gianollo* (University of Bologna)  
*Matti Miestamo* (University of Helsinki)  
*Marianne Mithun* (University of California Santa Barbara)

### Scientific Board

*Giorgio Francesco Arcodia* (Università Ca' Foscari, Venice)  
*Peter Arkadiev* (Johannes-Gutenberg University of Mainz)  
*Gilles Authier* (École Pratique des Hautes Études, Paris)  
*Luisa Brucale* (University of Palermo)  
*Holger Diessel* (University of Jena)  
*Eitan Grossman* (The Hebrew University of Jerusalem)  
*Corinna Handschuh* (Universität Regensburg)  
*Guglielmo Inglese* (University of Turin)  
*Elisabetta Magni* (University of Bologna)  
*Francesca Masini* (University of Bologna)  
*Susanne Maria Michaelis* (MPI EVA – Leipzig)  
*Emanuele Miola* (University of Bologna)  
*Anna Riccio* (University of Foggia)  
*Eva van Lier* (University of Amsterdam)

### Responsible Editor

*Caterina Mauri*, University of Bologna

Department of Modern Languages, Literatures and Cultures, via Cartoleria 5, 40124 Bologna. Email: [caterina.mauri@unibo.it](mailto:caterina.mauri@unibo.it)

### Production editors

*Silvia Ballarè* (University of Bologna)  
*Eleonora Zucchini* (University of Bologna)

### Assistant production editors

*Antonio Bianco* (University of Pavia)  
*Valentina Di Falco* (Independent assistant editor)  
*Sara Gemelli* (University of Pavia)  
*Maria Cristina Lo Baido* (University of Cagliari)  
*Nicola Perugini* (University of Bologna)  
*Antonia Russo* (University of Bergamo)  
*Federica Mauri* (Independent assistant editor)



# Linguistic Typology at the Crossroads



*Ideophones: honing in on a descriptive and typological concept*

Edited by

Aimée Lahaussois, Julie Marsault and Yvonne Treis

## CONTENTS

<b>Ideophones: honing in on a descriptive and typological concept - Editorial paper</b> <i>Julie Marsault, Aimée Lahaussois, Yvonne Treis</i> -----	1-13
<b>Terminological diversity in descriptions of Kiranti ideophonic lexemes</b> <i>Aimée Lahaussois</i> -----	14-43
<b>The sound roots of Umó<sup>n</sup>ho<sup>n</sup></b> <i>Julie Marsault</i> -----	44-90
<b>Teko ideophones: description of a word class</b> <i>Françoise Rose</i> -----	91-146
<b>Ideophones in Kambaata (Cushitic): Grammar, meaning and use</b> <i>Yvonne Treis</i> -----	147-200
<b>Ideophones and verbal constructions with the verb 'say' in Amharic</b> <i>Ronny Meyer</i> -----	201-243



# Linguistic Typology

---

## at the Crossroads



### **Ideophones in Sena (Bantu, Mozambique)**

*Rozenn Guérois* ----- 244 - 288

### **Ideophones and sound symbolism in Northern Amis (Austronesian)**

*Isabelle Bril* ----- 289 - 324

### **Ideophonic verb compounds in Archi**

*Gilles Authier* ----- 325 - 380

### **Ideophones in Upper Guinea Creoles: a comparative study**

*Nicolas Quint, Noël Bernard Biagui*----- 381 - 444



# Ideophones: honing in on a descriptive and typological concept

JULIE MARSAULT<sup>1</sup>, AIMÉE LAHAUSSOIS<sup>1,2</sup>, YVONNE TREIS<sup>3</sup>

<sup>1</sup>CNRS, UMR 7597 HISTOIRE DES THÉORIES LINGUISTIQUES

<sup>2</sup>UNIVERSITÉ PARIS CITÉ & UNIVERSITÉ SORBONNE NOUVELLE

<sup>3</sup>CNRS, UMR 8135 LANGAGE, LANGUES ET CULTURES D'AFRIQUE

Submitted: 9/04/2024    Revised version: 14/05/2024

Accepted: 14/05/2024    Published: 12/06/2024



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

## Abstract

This paper introduces the special issue of *Linguistic Typology at the Crossroads*, entitled “Ideophones: honing in on a descriptive and typological concept”, edited by Aimée Lahaussais, Julie Marsault and Yvonne Treis. The contributions, which are the result of work by a research group on ideophones and interjections, funded by the Labex EFL in Paris, are mainly descriptive papers with a typological perspective, informed by the more than 20 years of ideophone research following the landmark publication edited by Voeltz & Kilian-Hatz in 2001. After introducing the field and briefly presenting each contribution, we discuss issues in ideophone research that the articles gathered here contribute to, namely whether or not ideophones make up a clear-cut word class in different languages, and how different contributors interpret their relationship with onomatopoeia and interjections. We also present an overview of the primary syntactic uses of ideophones in the different languages and highlight the morphological processes attested for ideophonization and deideophonization, which are still under-investigated.

**Keywords:** ideophones; word classes; descriptive linguistics; typology; onomatopoeia; historiography.

## 1. Introduction

Ideophones have been a topic of research for more than a century, since even before the term “ideophone” was popularized by Doke (1935). Throughout the 20<sup>th</sup> century, ideophones have been the object of a growing number of descriptive works in individual languages, in linguistic families or in linguistic areas. While the term “ideophone” emerged in the Africanist tradition and is now the most widely used term, the Japonic linguistic tradition tends to use the term “mimetics” (e.g. Hirose 1981), while South Asian linguistics favors “expressives” (e.g. Diffloth 1976; Badenoch & Choksi 2021). For a comprehensive account of the history of research on ideophones, as well as a history of the term “ideophone”, see Dingemanse (2011a: section 2.1 and chapter 3).

Ideophones have been a topic of particular theoretical and typological interest since the turn of the 21<sup>st</sup> century (e.g. Hinton et al. 1994; Voeltz & Kilian-Hatz 2001; Akita 2009; Dingemanse 2012; Dingemanse 2017; Dingemanse 2019; McLean 2021; Andrason & Heine 2023; among many others), with recent edited volumes and monographs including Barrett et al. (2014), Armoskaite & Koskinen (2017), Haiman (2018), Akita & Pardeshi (2019), Badenoch & Choksi (2021). They have been shaped by interactions with research on sound symbolism and iconicity (e.g. Hinton et al. 1994; Nuckolls 1999; Dingemanse 2011b; Dingemanse & Akita 2016; Thompson & Do 2019; Winter et al. forthcoming).

Attempts at defining “ideophones” as a comparative concept for typological studies go back to the 2010s at least (esp. Dingemanse 2012; Dingemanse 2019; Akita & Pardeshi 2019). The cross-linguistic definition of “ideophones” proposed by Dingemanse (2012, 2019) is a welcome tool for typological work on ideophones, as proven by its nearly ubiquitous presence in articles on the topic (most of our contributions either use it or mention it), as well as for identifying diachronic variations, as he convincingly argues. The initial definition in 2012 was amended in 2019 to “[a] member of an open lexical class of marked words that depict sensory imagery” (Dingemanse 2019: 16). Dingemanse explicitly proposes this five-part definition as the basis for a comparative concept. Furthermore, the implicational hierarchy for the sensory domains depicted by ideophones, as proposed in Dingemanse (2012), has proven itself as a stimulating heuristic tool, as linguists have examined their own data with an eye towards confirming (or disproving) its predictions, which sometimes led to additions to or to minor revisions of the hierarchy

(e.g. McLean 2021, and comments in Rose this volume; for a very recent transformation of Dingemanse's hierarchy into a semantic map, see Van Hoey 2023).

One question which is not often addressed in the literature is the extent to which the category of “ideophones” is useful for the description of particular languages, language families, or areas. The papers collected in this special issue present work carried out by a research group funded by the Labex EFL, Paris (<https://en.labex-efl.fr/>) on “Ideophones and Interjections in a typological, areal and diachronic perspective”, bringing together researchers in the Paris area and some additional collaborators. In contrast to edited volumes which focus on specific features of ideophones or related theoretical issues (Lahti et al. 2014; Armoskaite & Koskinen 2017; Akita & Pardeshi 2019; Badenoch & Choksi 2021), the contributions in the present volume provide typologically-oriented descriptions of ideophones or ideophone-like words in under-described languages which are geographically and phylogenetically varied. In this sense, the volume has a similar approach to Voeltz & Kilian-Hatz (2001), but is informed by insights in the typological literature on ideophones of the last 20 years. The result is that the contributions to this volume describe a wider range of features of ideophones than had been previously taken into consideration. Additionally, the contributions for the most part include discussions on the definition and identification of ideophones, and on their status as a word class in individual languages. For instance, while the identification of a category of ideophones is straightforward in some languages, in others it presents a methodological challenge. In this way, the contributions gathered here participate in defining the boundaries between the “ideophone” as a word-class relevant to grammatical description and the set of iconic words that can be taken as reflecting ideophone-like phenomena but not constituting a definable class in a given language (Dingemanse 2019).

This volume explores ideophones in languages and linguistic areas that have been less well-represented in ideophone research, such as the Caucasus (Authier), North America (Marsault) and Creole languages (Quint), in most cases presenting new and unpublished field data. New aspects of ideophone research presented in this volume include the prosody of ideophones (Rose); the history of the use of the term “ideophone” and related terms in a linguistic area (Lahaussais); the presence of consonant or vowel gradation in ideophones or ideophone-like classes of words (Marsault, Rose, Bril), and deideophonic and ideophonicizing derivational processes (Meyer, Treis, Guérois). See also the discussion in section 3.



## 2. List of contributions

**Aimée Lahaussais** looks at ideophones in descriptions of **Kiranti languages** (Trans-Himalayan/Sino-Tibetan, Nepal), identifying them on the basis of morpho-phonological templates she has defined for languages on which she has carried out extensive fieldwork. Using these templates, she finds ideophonic lexemes in grammatical descriptions of other languages, and documents the associated terminology and descriptive apparatus. The study is motivated by the small footprint of ideophones in existing descriptions, and the possibility of carrying out areal work on these lexemes and exploring questions of language contact (and notably whether it is the patterns or the matter which are borrowed) if they were more easily identified as such in descriptions.

**Julie Marsault** analyzes sound roots in **Umó<sup>n</sup>ho<sup>n</sup>**, a Siouan language of North America, with a philological study based on a closed corpus. Adopting a top-down approach, she tries to determine whether Umó<sup>n</sup>ho<sup>n</sup> has a word class of ideophones that corresponds to Dingemans's comparative concept. She takes as a point of departure a nineteenth century publication listing many so-called "onomatopes" and identifies a group of sound roots with a coherent set of morphological, syntactic and semantic features. She concludes that the sound roots have ideophone-like features, but that there is no basis for identifying a distinct morphosyntactic class of ideophones.

**Françoise Rose** proposes an in-depth description of **Teko** (Tupi, French Guiana) ideophones entirely based on her own field data. In this language, ideophones form a clear-cut, easily identifiable category, and Rose analyzes their features systematically at various levels and compares them to the classes of nouns and verbs. This paper stands out for including a detailed analysis of prosodic features and discourse uses, as well as for providing quantitative data.

In **Kambaata** (Cushitic, Ethiopia) ideophones are an open word class of morphologically invariant lexemes that require the support verbs *y-* 'say' in intransitive clauses and *a'*- 'do' in transitive clauses to be inflected and syntactically integrated into an utterance. Based on lexicographical data and her own fieldnotes, **Yvonne Treis** discusses features of the phonology, phonotactics and stress marking of ideophones, investigates which derivational processes take ideophonic inputs, and shows in which syntactic functions ideophones are used.

**Amharic** (Ethiosemitic, Ethiopia) has complex predicates which consist of an invariant coverb followed by an inflecting light verb. **Ronny Meyer**'s article provides a detailed description of the elements that can function as coverbs in complex predicates and provides a language-internal morphosyntactic definition of ideophones as a word class. It is argued that only ideophonic coverbs can form noncausal/causal complex predicate pairs with the light verbs *alä* 'say' and *adärrägä* 'do'. In addition to a large number of basic ideophonic lexemes, Amharic can also derive ideophones from verbal roots through dedicated templates (non-linear morphology) expressing intensification or attenuation of the event denoted by the root.

Based on a recently collected fieldwork corpus, **Rozenn Guérois**'s paper offers an overview of ideophones in **Sena**, a Bantu language spoken in central Mozambique. By closely examining the different features (phonology, morphology, syntax and semantics) of Sena ideophones, her paper aims to identify the language-internal aspects of ideophones. The claim is that Sena ideophones are best treated as a distinctive word class whose members can be used in a variety of syntactic constructions. A crucial syntactic property is their ability to be used as independent and holophrastic predicates.

**Isabelle Brill** presents a study of ideophones in **Northern Amis** (East Formosan, Taiwan). Starting from Proto-Austronesian (PAn) "onomatopoeic" roots reconstructed by Blust (1988), she documents Amis words associated with sensory experiences, covering depictive to descriptive functions, and existing along a continuum with onomatopoeia on one end (as fully depictive lexemes) and fully grammatically integrated ideophones on the other (as fully descriptive lexemes). Questions regarding the status of ideophones are particularly interesting in light of the acategorical nature of lexemes in Northern Amis.

**Gilles Authier** describes the ideophones of **Archi** (East Caucasian, Dagestan/Russia), which are defined by their occurrence in light verb constructions with the verb *bos* 'say'. Based on data drawn from an online dictionary (Chumakina et al. 2007), Authier proposes a semantic classification of ideophones, which include sound and speech, non-auditory sensations, ingestion, movement and effortful activities, and places Archi ideophones within the larger context of language contact in the area.

**Nicolas Quint** investigates ideophones in **Upper Guinea Creoles** (West Africa), a family of Afro-Portuguese creoles, based on a combination of field data and published documentation. He investigates the inventory of ideophones in all varieties and shows

a divide between the continental branch and the insular (Cape Verdean) branch, which are mainly in contact with Niger-Congo languages and with Portuguese, respectively. The presence or absence of ideophones in these contact languages has had an impact on the size of the inventory and on the phonological and morphological features of ideophones in each branch of Upper Guinea Creoles.

### 3. Discussion

As mentioned above, one of our initial questions was about the word class status of ideophones and its language-internal usefulness as a descriptive concept. For the languages discussed in almost all the contributions, ideophones clearly constitute a distinct lexical category, based on morphological and/or syntactic criteria. In such cases, the authors can easily and successfully adopt a bottom-up approach to identify ideophones in the language, and then map them onto the comparative concept proposed by Dingemans (see esp. Treis, Rose, Meyer, Guérois). Other contributions show that although the comparative concept is not a perfect match for their data, the definition of ideophone is nonetheless useful and allows them to capture relationships between word types that might otherwise be overlooked (this is the case for Kiranti languages). A notable exception is the contribution by Marsault, who adopts a top-down approach to determine how well the comparative concept of ideophones fits the available Umó<sup>n</sup>ho<sup>n</sup> data, a methodology constrained by sound-imitating verbal roots which cannot be distinguished from other verbal roots in a systematic way.

Connected to the issue of word class status are questions about the boundary with other word types which tend to be grouped together under hypernyms such as “interactives” (Heine 2023) or “expressives” (Foolen 2015; “expressive” is particularly favored in discussing languages of South and Southeast Asia, see Badenoch & Choksi 2021; Williams 2021; Diffloth 1976). In the contributions to this volume, this especially concerns onomatopoeia and interjections.

**Ideophones vs. onomatopoeia.** Onomatopoeia are sometimes treated as a subtype of ideophones, as by Treis in this volume (see also Treis & Deginet 2024). This approach is also adopted in Haiman (2018: 82), Hamano (1998) for Japanese, and several contributions in Körtvélyessy & Štekauer (2024: 4). By contrast, Quint and Meyer specify that although sound-denoting ideophones are onomatopoeia, not all onomatopoeia are ideophones. Brill (section 2.2) states that onomatopoeia and ideophones are best described as two distinct categories in Northern Amis, although

a few lexemes belong to both categories (i.e., are sometimes used as onomatopoeia and sometimes as ideophones). This suggests that ideophones and onomatopoeia form two distinct but overlapping classes in these languages, in line with Dingemanse (2019: 28). Marsault, for her part, shows that the sound-imitative roots in Umó<sup>n</sup>ho<sup>n</sup> are completely distinct from onomatopoeia. The different analyses reveal a longstanding difficulty in defining onomatopoeia in typological/comparative terms, due to different definitions and uses of the term, and also due to important cross-linguistic variations concerning their relationship to other word classes. See Körtvélyessy & Štekauer (2024) for a recent handbook on onomatopoeia.

**Ideophones vs. interjections.** In contrast with onomatopoeia, all contributors who mention interjections treat them as a distinct word class from ideophones. This distinction is representative of the literature on the subject. Although ideophones and interjections are often cited together as typical representatives of the expressive or depictive function of language (e.g. Kunene 2001: 189), or for both having irregular phonic patterns (e.g. Ameka 2001: 30; Karani & Andrason 2022: 451), many descriptions make clear that they belong to distinct categories (e.g. Creissels 2001: 75-76; Dingemanse 2011a: 155-157; Karani & Andrason 2022: 455). Some of our contributions, however, mention interjections as adopting ideophonic properties in a specific syntactic environment. In Amharic and Kambaata (Meyer, Treis), ideophones only occur with the light verbs ‘say’ or ‘do’, in what Güldemann (2008) calls quotative constructions, and some interjections undergo semantic shifts when used with ‘say’, leaving them undifferentiated from ideophones. Quotative constructions thus constitute pivot constructions where the same lexeme can be used as an interjection or an ideophone (with distinct semantic results, only the first one being interpretable as reported speech). Authier briefly mentions a couple of conative interjections that can be used in quotative construction with a meaning distinct from direct reported speech (see also Reiter 2011: 478 ff. for interjections used with non-quotative light verbs). To the best of our knowledge, pivot constructions between ideophones and interjections have not been the object of dedicated studies yet.

Our contributions exemplify different systems of ideophones with regard to the ways they are integrated (or not) into the grammar. At the functional and semantic levels, most of the languages described have ideophones with either primarily predicative uses or primarily adverbial uses. This kind of distinction is already widely commented on in the literature, including language-internally. For instance, Güldemann (2008: 282–283) notes that most ideophones, depending on the language,

either occur in regular collocation with another content word to intensify its meaning, or they are semantically and structurally independent. In the second use, he distinguishes between ideophones that “establish [...] an event representation on their own” (2008: 282) and “inserting constructions” (2008: 282) with a ‘say’-verb. For his part, and in parallel to Güldemann, Heine (2023: 153) identifies modifying, free and quotative constructions of ideophones.

We can cite for **primary predicative uses** Teko (Rose), Kambaata (Treis), Amharic (Meyer), Archi (Authier) and Northern Amis (Bril). All of these languages have quotative constructions. Quotative constructions form the only possible syntactic frame for ideophones in Kambaata, Amharic and Archi (and the authors use this as the main criterion for identifying ideophones), while Teko ideophones can also be used as (argument-taking) predicates on their own, and Northern Amis ideophones can be used as (argument-taking) predicates when they take voice markers. By contrast, ideophones in Upper Guinea Creoles (Quint) and the Kiranti languages (Lahaussais) have **adverbial uses**, where they cover different semantic fields but always modify a verb phrase or an adjective. Note that both families have a subcategory of ideophones with an intensifying meaning and that generally combine with only one lexeme. Intensifying ideophones are well-known in the literature (e.g. Newman 1968: 109; Bowler & Gluckman 2018). In Sena (Guérois), ideophones can be used predicatively or adverbially. In their predicative use, they can be used as predicates on their own, or in a complex predicate with a light verb, in which case they are semantically equivalent to nouns or adjectives (in attributive predicate functions). The sound-denoting roots in Umó<sup>n</sup>ho<sup>n</sup> (Marsault) are verbal roots, but they are often used in a verb-modifying function in verb series, as Marsault demonstrates.

Finally, both Teko and Sena also feature ideophones in **holophrastic uses**, which are close to predicative uses (we include in this category Rose’s “collocational use” of Teko ideophones, where the ideophone forms its own clause but co-expresses an event expressed by a verb in the following clause). In summary, except for the independent uses of ideophones attested in Teko and Sena, all the examples in our contributions show relatively well integrated ideophones, syntactically speaking. (For a recent publication addressing the ambivalent grammatical behavior of ideophones as syntactically detached interactive elements and as fully morphosyntactically integrated constituents, see Andrason & Heine 2023, who view these (and other) uses as different stages in a gradual grammaticalization process).

A still underrepresented issue in the literature is the issue of derivation into or out of the category of ideophones – although note works on morphological ideophonization processes such as Childs (1989) on Kisi and Le Guen (2014) on Yucatec Maya; Jacques (2013) describes both word class-preserving ideophonic morphology as well as the formation of deideophonic verbs and nouns. Most contributions to this special issue mention cases of **deideophonization through derivation** (and not merely through syntactic integration and/or semantic shift, resulting in deideophonic verbs, nouns, or adjectives).<sup>1</sup> Notably, Treis identifies two productive Kambaata suffixes dedicated to deriving resultative adjectives and action nouns from ideophones respectively. Brill discusses the formation of deideophonic nouns, verbs and modifiers in Northern Amis. Quint shows that a verbalizing suffix is attested on a few intensifying ideophones of Casamance Upper Guinea Creole, and the resulting verb is a semantic equivalent to the collocation the ideophone originally occurred in. By contrast, in Archi it is the whole compound <ideophone + ‘say’> that is turned into an adjective. In the opposite direction, Guérois and Meyer mention productive **morphological processes of ideophonization**. In the case of Sena, intensifying ideophones are derived from verbs and are used in collocation with them. In the case of Amharic, a language with non-linear morphology, two templates for trilateral roots are used to create ideophones with intensive and attenuative readings from the corresponding verbs (and are used predicatively with a light verb). These templates are very productive, as Meyer notes. Moreover, they do not seem to be inherited from proto-Semitic or proto-Ethiosemitic.

As for the individual details surrounding ideophones in these different languages, readers are invited to refer to the articles collected here!

## **Acknowledgements**

This work is supported by a public grant overseen by the IdEx Université Paris Cité (ANR-18-IDEX-0001) as part of the LabEx *Empirical Foundations of Linguistics* (EFL).

---

<sup>1</sup> We are aware that the distinction between derivation and inflection (the latter being analyzable as a consequence of conversion or zero-derivation) can be tricky. Borderline examples in our contributions include Archi’s ideophonic compounds (ideophone + verb ‘say’) turned into adjectives via participle marking.

## References

- Akita, Kimi. 2009. *A grammar of sound-symbolic words in Japanese: Theoretical approaches to iconic and lexical properties of Japanese mimetics*. Doctoral dissertation, Kobe University.
- Akita, Kimi & Prashant Pardeshi (eds.). 2019. *Ideophones, mimetics and expressives*. Amsterdam: John Benjamins. <https://doi.org/10.1075/ill.16>
- Ameka, Felix K. 2001. Ideophones and the nature of the adjective word class in Ewe. In F. K. Erhard Voeltz & Christa Kilian-Hatz (eds.), *Ideophones*, 25–48. Amsterdam: John Benjamins.
- Andrason, Alexander & Bernd Heine. 2023. On the grammaticalization of ideophones. In Sylvia Hancil & Vittorio Tantucci (eds.), *Different slants on grammaticalization*, 237–262. Amsterdam: John Benjamins.
- Armoskaite, Solveiga & Päivi Koskinen (eds.). 2017. Structuring sensory imagery: Ideophones across languages and cultures. Special issue of *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 62(2). <http://dx.doi.org/10.1017/cnj.2017.12>
- Badenoch, Nathan & Nishaant Choksi (eds.). 2021. *Expressives in the South Asian Linguistic Area*. Leiden: Brill.
- Barrett, Rusty & Anthony K. Webster & Tuomas Huumo. 2014. Introduction: Ideophones: Between grammar and poetry. *Pragmatics and Society* 5(3). 335–340.
- Blust, Robert. 1988. *Austronesian Root Theory; an essay on the limits of morphology*. Amsterdam: John Benjamins.
- Bowler, Margit & John Gluckman. 2018. Intensifying ideophones in three Luhya languages. In Elizabeth Bogal-Allbritten & Elizabeth Coppock (eds.), *Proceedings of TripleA 4: Fieldwork perspectives on the semantics of African, Asian and Austronesian languages, University of Gothenburg, June 9-11, 2017*, 31–47. Tübingen: Universitätsbibliothek Tübingen.
- Childs, G. Tucker. 1989. Where do ideophones come from? *Studies in the Linguistic Sciences* 19(2). 55–76.
- Chumakina, Marina & Dunstan Brown & Greville G. Corbett & Harley Quilliam. 2007. *A dictionary of Archi: Archi-Russian-English* (Online edition). University of Surrey. <http://dx.doi.org/10.15126/SMG.16/2>

- Creissels, Denis. 2001. Setswana ideophones as uninflected predicative lexemes. In F. K. Erhard Voeltz & Christa Kilian-Hatz (eds.), *Ideophones*, 75–86. Amsterdam: John Benjamins.
- Diffloth, Gérard. 1976. Expressives in Semai. In *Austroasiatic Studies Part 1* (Oceanic Linguistics Special Publications 13), 249–264. Honolulu: University of Hawai'i Press.
- Dingemanse, Mark. 2011a. *The meaning and use of ideophones in Siwu*. Doctoral dissertation, Radboud University, Nijmegen.
- Dingemanse, Mark. 2011b. Ezra Pound among the Mawu: Ideophones and iconicity in Siwu. In Pascal Michelucci & Olga Fischer & Christina Ljungberg (eds.), *Semblance and signification*, 39–54. Amsterdam: John Benjamins.
- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6(10). 654–672.  
<https://doi.org/10.1002/lnc3.361>
- Dingemanse, Mark. 2017. Expressiveness and system integration: On the typology of ideophones, with special reference to Siwu. *Language Typology and Universals* 70(2). 363–384. <https://doi.org/10.1515/stuf-2017-0018>
- Dingemanse, Mark. 2019. 'Ideophone' as a comparative concept. In Kimi Akita & Parshant Pardeshi (eds.), *Ideophones, mimetics and expressives*, 13–34. Amsterdam: John Benjamins.
- Dingemanse, Mark & Kimi Akita. 2016. An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics* 53(3). 501–532.  
<http://dx.doi.org/10.1017/S002222671600030X>
- Doke, Clement M. 1935. *Bantu linguistic terminology*. London: Longmans Green.
- Foolen, Ad. 2015. Expressives. In Rick Riemer (ed.), *The Routledge handbook of semantics*, 473–490. London: Routledge.
- Güldemann, Tom. 2008. *Quotative indexes in African languages*. Berlin: De Gruyter Mouton.
- Haiman, John. 2018. *Ideophones and the evolution of language*. Cambridge: Cambridge University Press.
- Hamano, Shoko. 1998. *The sound-symbolic system of Japanese*. Stanford, CA: CSLI Publications.
- Heine, Bernd. 2023. *The grammar of interactives*. Oxford: Oxford University Press.



- Hinton, Leanne & Johanna Nichols & John J. Ohala (eds.). 1994. *Sound symbolism*. Cambridge: Cambridge University Press.
- Hirose, Masayoshi. 1981. *Japanese and English contrastive lexicology: The role of Japanese “mimetic adverbs”*. Doctoral dissertation, University of California at Berkeley.
- Jacques, Guillaume. 2013. Ideophones in Japhug (Rgyalrong). *Anthropological Linguistics* 55(3). 256–287.
- Karani, Michael & Alexander Andrason. 2022. Ideophones in Arusa Maasai: Syntax, morphology, and phonetics. *Open Linguistics* 8(1). 440–458. <https://doi.org/10.1515/opli-2022-0220>
- Körtvélyessy, Lívia & Pavol Štekauer (eds.). 2024. *Onomatopoeia in the world’s languages*. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783111053226>
- Kunene, Daniel P. 2001. Speaking the act. In F. K. Erhard Voeltz & Christa Kilian-Hatz (eds.), *Ideophones*, 183–191. Amsterdam: John Benjamins.
- Lahti, Katherine & Rusty Barrett & Anthony K. Webster (eds.). 2014. *Ideophones: Between grammar and poetry* (Special issue of *Pragmatics and Society* 5.3). <https://doi.org/10.1075/ps.5.3>
- Le Guen, Olivier. 2014. Expressive morphology in Yucatec Maya. In Jean Léo Léonard & Alain Kihm (eds.), *Patterns in Mesoamerican morphology*, 178–211. Paris: Michel Houdiard.
- McLean, Bonnie. 2021. Revising an implicational hierarchy for the meanings of ideophones, with special reference to Japonic. *Linguistic Typology* 25(3). 507–549. <https://doi.org/10.1515/lingty-2020-2063>
- Newman, Paul. 1968. Ideophones from a syntactic point of view. *Journal of West African languages* 5. 107–117.
- Nuckolls, Janis B. 1999. The case for sound symbolism. *Annual Review of Anthropology* 28. 225–252.
- Reiter, Sabine. 2011. *Ideophones in Awetí*. Doctoral dissertation, Universität zu Kiel.
- Thompson, Arthur Lewis & Youngah Do. 2019. Defining iconicity: An articulation-based methodology for explaining the phonological structure of ideophones. *Glossa: a journal of general linguistics* 4(1):72. <https://doi.org/10.5334/gjgl.872>
- Treis, Yvonne & Deginet Wotango Doyiso. 2024. Onomatopoeia in Kambaata. In Lívia Körtvélyessy & Pavol Štekauer (eds.), *Onomatopoeia in the world’s languages*, 81–92. Berlin: De Gruyter Mouton.

- Van Hoey, Thomas. 2023. A semantic map for ideophones. In Thomas Fuyin Li (ed.), *Handbook of cognitive semantics*, vol. 2, 129–175. Leiden: Brill.  
<https://doi.org/10.31219/osf.io/muhpd>
- Voeltz, F. K. Erhard & Christa Kilian-Hatz (eds.). 2001. *Ideophones*. Amsterdam: John Benjamins.
- Williams, Jeffrey P. (ed.). 2021. *Expressive morphology in the languages of South Asia*. Oxon: Routledge.
- Winter, Bodo & Greg Woodin & Marcus Perlman. forthcoming. Defining iconicity for the cognitive sciences. In Olga Fisher & Kimi Akita & Pamela Perniss (eds.), *Oxford handbook of iconicity in language*. Oxford: Oxford University Press.

**CONTACT**

julie.marsault@inalco.fr  
aimee.lahaussois@cnrs.fr  
yvonne.treis@cnrs.fr

# Terminological diversity in descriptions of Kiranti ideophonic lexemes

AIMÉE LAHAUSSOIS

UNIVERSITÉ PARIS CITÉ & UNIVERSITÉ SORBONNE NOUVELLE  
CNRS, UMR 7597 HISTOIRE DES THÉORIES LINGUISTIQUES

Submitted: 30/01/2023    Revised version: 23/08/2023

Accepted: 09/10/2023    Published: 12/06/2024



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

## Abstract

In this article, I document terms which have been used for ideophonic lexemes in descriptions of Kiranti (Trans-Himalayan/Sino-Tibetan) languages and their evolution over time. This involves identifying ideophonic lexemes in existing descriptions of these languages, on the basis of morphophonological patterns they tend to share throughout the Kiranti area (Eastern Nepal), collecting all the associated terms, and placing them in the context of the descriptions in which they appear. This analysis of the terminological choices by different authors in descriptive work on these languages is motivated by a desire to make data on ideophones in this linguistic area more accessible to a typologically oriented readership. In so doing, it raises awareness about the difficulties in describing these lexemes before the appearance of a unified comparable framework, something we now have access to through work by researchers such as Dingemans.

**Keywords:** Nepal; Kiranti grammaticography; ideophones; onomatopoeia; expressives.

## 1. Introduction

Ideophones are an increasingly popular research topic for South Asian languages, judging from the growing number of publications presenting descriptions of

ideophones in individual languages. Across the linguistic area, scholars (Emeneau 1969; Masica 1976; Dahal 1999; Pokharel 1993; Abbi 2018) have long noted the presence of words, often marked by reduplication, that depict sound as well as other sensory inputs, and such ideophone-like words are arguably found in many, if not all, Kiranti languages (Eastern Nepal, Trans-Himalayan/Sino-Tibetan). Their properties are most extensively described in dedicated articles (Rai & Winter 1997; Rai et al. 2005; Lahaussais 2017a, 2023), but they are also present, albeit couched in a variety of terms, in many of the descriptive grammars of languages of the subgroup.

One of the major problems faced in investigating similarities among ideophones across languages is access to material that can be compared. The terminology for ideophone-like lexemes is quite diverse across Western language descriptions of Kiranti languages: among the labels for material that conforms to morphophonological templates assumed to be ideophonic (Lahaussais 2023; for more on the templates, see § 2), we find terms such as “onomatopoeia”, “expressive”, “paralexeme”, “ideophone”, “phonesthetic word”, “mimetic”, “adverb”. The result of this terminological profusion is that cross-linguistic work on ideophonic lexemes may be hampered by difficulties in identifying the relevant materials in descriptions. This of course applies to terminology in all subdomains of linguistics, and is behind efforts concerning the standardization of terminological choices (see, e.g., Chelliah, Burke & Heaton 2021) which are important to our ability to integrate languages of this area (and elsewhere) into typological research.

This article will focus on the terminological diversity found in the description of ideophones in Kiranti languages, but the range of terms will be familiar to readers working on Trans-Himalayan<sup>1</sup> languages and beyond. The ultimate goal of this contribution is to provide insight into the distribution of terms, the chronology of their use, and to shed light on the potential use of the materials for larger-scale comparison.

Section 2 presents the sources and data used in the study; Section 3 presents the terminology found across the descriptions which discuss ideophones (or lexemes sharing ideophonic characteristics); Section 4 is a discussion of the relevance of the

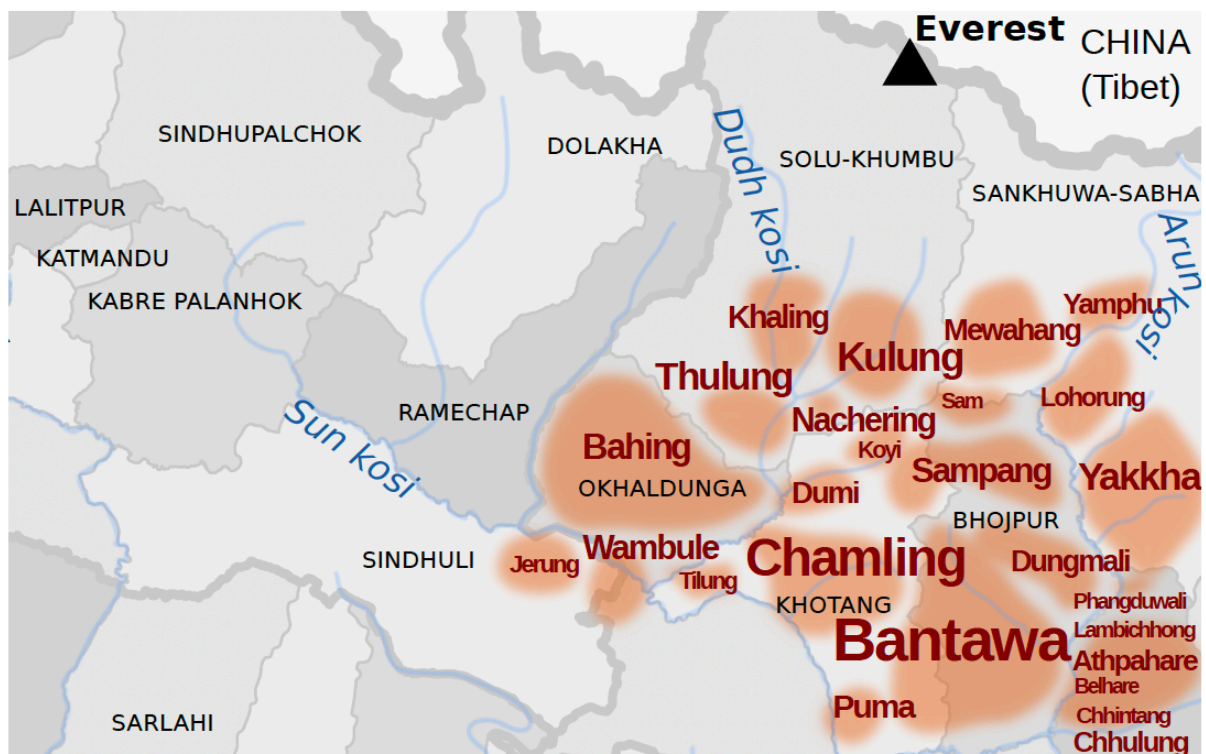
---

<sup>1</sup> This family has been known as Sino-Tibetan and Tibeto-Burman at various times in its past; Trans-Himalayan is considered a more neutral term, intended as a purely geographical label which makes no claims about higher-order grouping (van Driem 2018).

terminological choices and of what is gained and lost by the choice of a particular term. Section 5 presents my conclusions.

## 2. Data and methods

The 30-odd Kiranti languages are spoken in Eastern Nepal, and generally considered to form a subgroup,<sup>2</sup> belonging to the Trans-Himalayan language family. Their geographical distribution is shown in Figure 1. For information on the general characteristics of Kiranti languages, see Michailovsky (2017) and Ebert (1994).



CC-BY-SA Sebastian Nordhoff

Figure 1: Map of the Kiranti area, adapted from Schlemmer (2021).

<sup>2</sup> See, however, Gerber & Grollman (2018) for a discussion of why this is not a given.

## 2.1. Database

The Western-metalanguage descriptive tradition for Kiranti languages dates back to the mid-19th century, beginning with two sketches by Hodgson of Vayu (VAY<sup>3</sup>; now usually Hayu or Wayu) and Bahing (BHJ), published in 1857 and 1857-58 respectively (Hodgson 1857a, 1857b, 1858). These are followed by the sketches in the *Linguistic Survey of India (LSI)*, with Kiranti materials appearing in Volume 3.1 (Grierson 1909). For the most part, the *LSI* materials are taken from pre-existing sources, such as Hodgson's aforementioned sketches, and comparative word and phrase lists, and supplemented by specimen texts, including the Parable of the Prodigal Son collected by Grierson's collaborators (Majeed 2019a, 2019b; Lahaussais 2021). Some thirty years after the *Linguistic Survey of India* sketches, sketches of Sangpang (RAV), Khambu (KHAM1300), Kulung (KLE) and Thulung (TDH) were produced by Wolfenden (1933a, 1933b, 1934, 1935) on the basis, again, of the Parable of the Prodigal Son, collected by Wolfenden himself in the four languages. None of the above materials describe or present anything resembling ideophones.

The grammaticography<sup>4</sup> of Kiranti languages picks up considerably in the 1970's with work on Khaling (KLR) (Toba 1984), Thulung (Allen 1975), Bantawa (BAP) (Rai 1984), Limbu (LIF, Weidert & Subba 1985; van Driem 1987), and Hayu (Michailovsky 1988). There follows a period of intensive activity in the description of these languages, through a large number of dissertations and published grammars: Dumi (DUS) (van Driem 1993; Rai 2016), Camling (RAB) (Ebert 1997a; Rai 2012), Athpare (APH) (Ebert 1997b), Yamphu (YBI) (Rutgers 1998), Wambule (WME) (Opgenort 2004), Jero (JEE) (Opgenort 2005), Kulung (KLE) (Tolsma 2006), Chhatthare Limbu (LIF) (Tumbahang 2007; Tumbahang 2017), Sunwar (SUZ) (Borchers 2008), Bantawa (BAP) (Doornenbal 2009), Puma (PUM) (Sharma 2014), Koyee (KKT) (Rai 2015), Chintang (CTN) (Paudyal 2015), Yakkha (YBH) (Schackow 2015).<sup>5</sup> The study presented in this article takes as its primary sources these 23 more recent grammatical descriptions, which are supplemented by four articles which focus on ideophones (see Table 3).

---

<sup>3</sup> All abbreviations for language names are ISO 639-3; dialects do not have their own code, and the language code is thus reprised for different dialects.

<sup>4</sup> The term, which is in common use in my research group on the history of linguistics and of linguistic description (*Histoire des théories linguistiques*, UMR 7597), is defined by Lehman & Maslova (2004: 1857) as follows: "At the object level, grammaticography is (the practice \_ experience or art \_ of) grammar writing. At the meta-level, it is the investigation of methodological principles that reconcile this practice with linguistic theorizing." I prefer to adopt the term "metagrammaticography" for the second, 'meta' definition, and to reserve "grammaticography" for grammar-writing.

<sup>5</sup> Ongoing work on Kiranti should soon result in grammars of Mewahang and of Nachiring.

The sketch descriptions found in Thurgood & LaPolla (2017), some of which are on Kiranti languages, are not taken into account here, nor are theses and dissertations focusing on a specific topic.<sup>6</sup>

## 2.2. Methodology

In this study, I seek to understand how the terms across the sources of the database differ and how they have spread. I do this not with any prescriptive ambitions, but in order to draw attention to the fact that many more Kiranti grammars contain ideophonic material than I had previously suspected (Kelly & Lahaussais 2021) and that the terms chosen by different grammarians may obscure this fact.

The methodology adopted here was to search the descriptions for lexemes conforming to morphophonological patterns identified as ideophonic in Kiranti (Lahaussais 2023) and to document any information relating to their description in the sources: adopted terminology, explanations of the choice of the term, comparisons with other existing terms, and any other patterns or templates associated with these other terms.

The morphophonological patterns I used to identify these lexemes are of four types. The very same or similar patterns are found across multiple Kiranti languages, and associated with an adverbial (and sometimes adjectival) function. The examples illustrating these patterns in (1) - (7) are all from my corpus of Thulung.<sup>7</sup>

a) A preverbal pattern, which tends to be monosyllabic, but can also be bisyllabic; when the latter, it tends strongly to have the same vowel in both syllables, and exists, in many of these cases, as a variant to the monosyllabic version. The ideophonic preverb is bolded in (1):

(1)	<i>bloku-ra</i>	<i>hoŋkorma</i>	<i>jok-ta-lo</i>	<i>mutstɯ</i>	<i>wo</i>
	river-LOC	flood	come.down-PST-TEMP	person	also
	<b><i>sep</i></b>	<i>bak-tɯ</i>			
	IDEO.PRVB	sweep.away-3SG > 3SG.PST			

---

<sup>6</sup> Bickel (1996) on TAM in Belhare is, however, used as a datapoint, as the first work to refer to ideophones by name.

<sup>7</sup> The corpus consists of more than 10 hours of annotated data, collected over twenty years; a considerable number of the texts are available on the Pangloss archive, [https://pangloss.cnrs.fr/corpus/Thulung\\_Rai?lang=en](https://pangloss.cnrs.fr/corpus/Thulung_Rai?lang=en)

‘In the river, when the flood came down, it swept away people as well.’ (Field notes, Dec. 2022)

Ideophonic preverbs of this type, in Thulung, have semantics which convey suddenness, abruptness, or thoroughness, and generally collocate with a limited set of action verbs. They cannot host derivational morphology and are inseparable from the collocating verb, with nothing able to be inserted between the two.

b) A reduplicated pattern which is not morphological (i.e. the unreduplicated form is not found). The reduplication can be full or partial. This pattern, illustrated in (2)-(4), includes (adverbial) onomatopoeia, as in (3).

(2) *khurukhuru*      *mi-dzəpa*      *lapdi*      *mi-lak-tsi*  
 IDEO.continuously    NEG-good      road    NEG-go-2DU  
 ‘Do not go continuously on the bad road.’ (Eagle story)

(3) *ḡhunḡkunḡhunḡkunḡ*      *ḡm-mu*  
 IDEO.drumming.sound      beat-INF  
 ‘to make a drum sound.’ (Field notes, Dec. 2022)

(4) *sunjaksunjak*      *khram-mu*  
 IDEO.soundlessly      cry-INF  
 ‘to cry soundlessly.’ (Field notes, Dec. 2022)

This pattern is able to host nominalizing morphology, in which case the lexeme can be used attributively (or referentially).<sup>8</sup>

c) A triplicated pattern, which results in a three-syllable lexeme; in some languages, the pattern can be fully triplicated, with the same syllable throughout, while in others, it is partially triplicated,<sup>9</sup> with one initial consonant in the first syllable and a different

<sup>8</sup> For a description of nominalization and its functions, see e.g. Lahaussais (2003; 2017b) for Thulung.

<sup>9</sup> In the consulted data, only Chintang has both partially and fully triplicated ideophones. Other languages seem to have one or the other type only.



initial consonant in the second and third but the same vowel throughout. Thulung only has partial triplication, illustrated in (5)–(6).

- (5) *par-laŋka*    *kaŋ*    ***plititi***    *luk-ta*  
wound-ABL    pus    IDEO.oozing    come.out-3SG.PST  
‘The pus oozed out of the wound.’ (Field notes, Dec. 2022)

- (6) *wakha:tam*    ***sututu***    *tsar-so-ra*  
slowly    IDEO.sliding    drop-DER-2SG > 3SG.IMP  
‘Drop it slowly in a sliding motion.’ (Field notes, Dec. 2022)

d) A pattern ending in a geminated consonant followed by  $\iota(i)$  or  $a(i)$ ; this pattern is well attested in Nepali, and the extent to which these words are loans is not clear across languages where it is found, despite some examples in Thulung which are not found in Nepali, such as that in (7).

- (7) *u-miksi-ra*                      *memsaka-ŋa*    *suk-tɕ*                      *ʔe*  
3SG.POSS-eye-LOC    like.that-INT    stick.in-3SG > 3SG.PST    HS  
***grwappai***                      *suk-tɕ*                      *ʔe*  
IDEO.forcefully    stick.in-3SG > 3SG.PST    HS

‘She stuck it into his eye like that, she **forcefully** stuck it in.’ (Eagle story)

### 2.3. Distribution of ideophones across the corpus

The same patterns have been found across the Kiranti languages, with the dominant patterns being the reduplicated, found in all descriptions, and the triplicated, found in half (7 of 14) of the languages whose descriptions were examined in Lahaussis (2023, Table 6).

Across the sources, we find no descriptions (not even in word lists) of ideophonic lexemes before the 1970’s. Once ideophones start appearing in grammars, it is with rather diverse terminology, as will be discussed in § 3, and with different levels of detail. Table 1 shows, for each grammar featuring sections that at least minimally discuss ideophones, how they are presented. Table 1 indicates the title of the section in question, the higher-level chapter that the section belongs to, and the approximate length of the

section on ideophones. Except where indicated, glosses for the ideophones are always lexical, and the transcriptions in all these sources are in IPA (or an adaptation of IPA, using *c* and *j* for affricates).

Language	Position within grammar (+ title)	Length	Presentation of examples
<b>Thulung</b> (Allen 1975)	§ 3.1 Adverbs of abruptness; 3.2 Reduplication; 3.3 With -maksī (all three sections are within chapter on Adverbs)	2 pages	lexical glosses or free translation of example with no glossing
<b>Bantawa</b> (N.K. Rai 1984)	§ 5.2.4 Onomatopoeic adverbs (within chapter on Adverbs)	1.5 pages	
<b>Hayu</b> (Michailovsky 1988)	§ 2.6.3 Mots phonéthétiques (within chapter on phonology)	1 page	free translation but no glossing
<b>Limbu</b> (Weidert & Subba 1985)	§ 4.5 Adverbs and expressives (within chapter on Morphology)	0.5 page	no examples
<b>Wambule</b> (Opgenort 2004)	§ 5.10 Adverbial proclitics of manner (within chapter on Nominals and adverbials); also in appended glossary, glossed as compound adverbs or adverbs.	1.5 pages	
<b>Chhatthare Limbu</b> (Tumbahang 2007 <sup>10</sup> )	§ 2.2.8 Derivation of adverbs by reduplication and prefixation (within chapter on Morphology of adverbs)	1 page	free translation but no glossing
<b>Bantawa</b> (Doornenbal 2009)	§ 8.2.4 Reduplication in adverbs (within chapter on Other word classes)	1.5 pages	
<b>Yakkha</b> (Schackow 2015)	§ 6.4 Reduplication, triplication and ideophones (within chapter on Adjectives and adverbs)	8.5 pages	

<sup>10</sup> The section dealing with ideophones is quasi-identical to the one (of the same name) in the later (2017) grammar by the same author, and the latter is therefore not entered into this table.

Language	Position within grammar (+ title)	Length	Presentation of examples
Koyee (T.M. Rai 2015)	§ 11.4 Onomatopoeia (within chapter on Adverbs and other minor word classes)	1 page	
Dumi (N.M. Rai 2016)	§ 10.2.7 Expressive adverbs (within chapter on Adverbs and postpositions)	1 page	

**Table 1:** Grammars with sections on ideophonic lexemes.

In some other grammars, the data on ideophones are not described in the main body of the grammar, but instead found in accompanying glossaries or in interlinearized examples (either in the text of the grammar or in appended texts), with the accompanying terminology is in the form of a gloss. The grammars of this type are listed in Table 2, along with the part of the grammar where ideophones are found and the glosses they receive.

Language	Position within grammar	Presentation of examples
Limbu (van Driem 1987)	glossary	gloss ‘onomatopoeia’
Dumi (van Driem 1993)	glossary	gloss ‘onomatopoeia’
Athpare (Ebert 1997)	glossed examples and texts	gloss ‘IDEO’
Camling (Ebert 1997, 2000 <sup>11</sup> )	glossed examples and texts	gloss ‘IDEO’
Yamphu (Rutgers 1998)	glossary	gloss ‘adv.onom.’, ‘adj.redup.’, ‘adv.redup.’, ‘adv.’
Puma (Sharma 2014)	a few scattered mentions	no examples, but text uses term ‘idiophones’ [sic]

**Table 2:** Grammars which mention ideophones but do not have dedicated sections.

<sup>11</sup> While Ebert (2000) is not a grammar, it is the volume of texts that serves as a companion to the 1997 grammar of Camling, and is thus considered here for the purposes of identifying mentions of ideophones.

The description of Belhare (BYW) by Bickel (1996) deserves mention, even though it is not a grammar per se and therefore not included in Table 2. It focuses on a specific feature, namely tense, aspect and mood in Belhare. Some examples, which are interlinearized, contain ideophonic lexemes, which are glossed ‘IDEOPH’. This is in fact the earliest instance of “ideophone” as a term in the database.

Another set of sources which were explored are dedicated articles, specifically focusing on ideophones. They are listed in Table 3, as they are presumably important in potentially shaping the descriptions of ideophones appearing in successive grammars.

Language	Main term	Subclasses
<b>Bantawa</b> (N.K.Rai & Winter 1997)	paralexeme	triplicated verbal adjunct <sup>12</sup>
<b>Chintang</b> (N.K.Rai et al. 2005)	ideophone	2 subclasses: reduplicated form; triplicated form <sup>13</sup>
<b>Khaling</b> (Lahaussais 2017a)	ideophone	3 subclasses (based on morphophonological pattern) <sup>14</sup>
<b>Thulung</b> (Lahaussais 2023)	ideophone	4 subclasses (based on morphophonological pattern)

**Table 3:** Dedicated articles on ideophones in Kiranti.

Other grammars, listed in § 2.1. as part of the corpus of Kiranti grammars, do not contain any material that I am able to identify as pertaining to ideophones: this is the case, in addition to the pre-1970’s grammars, for descriptions of Khaling (Toba 1984); Jero (Opgenort 2005), Kulung (Tolsma 2006), Sunwar (Borchers 2008) and Chamling (Rai 2012), regardless of the terminology.

### 3. Terminology found in the corpus

The material presented in Tables 1, 2, and 3 gives us a set of terms associated with ideophones across the grammars of the corpus:

<sup>12</sup> This subclass is related (but not identical) to Thulung type c illustrated in § 2.2.

<sup>13</sup> These subclasses are related (but not identical) to Thulung type b and c respectively, illustrated in § 2.2.

<sup>14</sup> These are the equivalents of Thulung subtypes a, b, and c.

adverb, reduplication, onomatopoeia/onomatopoetic, phonesthetic, expressive, adverbial proclitics, triplication, ideophone, paralexeme<sup>15</sup>

Note that some of the terms are nouns and other adjectives. One consequence of the presence of adjectival forms is that “adverb”, treated in § 3.6., can occur in combination with the other, more specialized terms from the list above.

Another point deserving mention is that other terms also surface in the grammars, often used as synonyms or comparanda for ideophonic materials and accompanying the main term adopted by the linguist; while other traditional parts of speech rarely get explained by means of alternative terms, ideophones do, suggesting that the main term is felt to be insufficiently well-defined to stand alone.

The subsections which follow explore the use in the individual grammars of the various labels for ideophonic lexemes. The lexemes covered by these terms map onto the morphophonological patterns presented for Thulung in § 2.2., regardless of the label they are assigned.

### **3.1. Expressive**

Three grammars in the database use the term “expressive”, usually as an adjective, although in one of the grammars, a nominal form is also found.

Allen, writing on Thulung, uses an umbrella category of “expressive resources” (Allen 1975: 107) for the various subclasses of adverbs which have ideophonic characteristics. The labels given to the three subclasses are the following: “adverbs of abruptness” (Allen 1975: 107), which correspond to the ideophonic preverbs described in § 2 above; “reduplicate adverbs and adjectives” (Allen 1975: 106), which correspond to the reduplicated pattern; and those corresponding to the triplicated pattern, which are not given an explicit label but appear under the heading “Reduplication” (Allen 1975: 108). Allen makes an explicit connection with ideophones in other languages and the Thulung material: “Like the ideophones of many African languages, the expressive adverbs seem to be regarded by native speakers as marginal to the language proper” (Allen 1975: 107).

---

<sup>15</sup> A reviewer wondered whether ‘verbal classifier’, matching the preverbal pattern exemplified for Thulung in § 2.2., was found; it appears in none of the grammars making up the database used here.

In the grammar of Limbu (Weidert & Subba 1985), the first occurrence of “expressive” in the grammar is accompanied by a paraphrase (“In the case of expressive or phonaesthetic adverbs” (Weidert & Subba 1985: 15)). It is said of “expressive adverbs” that they “are not exactly a homogeneous word class because there is no coherent marking system. But many of them are easily recognized due to the repetition or near-repetition of the verb syllable with which they get connected morphosyntactically” (Weidert & Subba 1985: 53-54). The Limbu dictionary which follows the grammar in the same volume presents quite a number of lexemes, glossed as expressives: some of them map onto the monosyllabic ideophonic preverbal pattern (illustrated for Thulung in (1) above); others show a reduplicated pattern.

In N.M. Rai’s (2016) grammar of Dumi, the language is described as having a subclass of adverbs called “expressive adverbs” (Rai 2016: 300). They include reduplicated and triplicated forms. Some of the reduplicated adverbs are said to have verbal roots, but others, including some triplicated adverbs, are said to have “onomatopoeic sources” (Rai 2016: 301). Other ideophonic lexemes are found under the label “onomatopoeic nouns” (Rai 2016: 178), although their glosses suggest an adverbial function.<sup>16</sup> This suggests that “expressive adverbs” are a word class corresponding to ideophones, of which a subclass is sound-imitative, for which the label “onomatopoeic” (see § 3.2.) is used.

The terms “expressive” as a modifier was popular among Prague School linguists (Diffloth 2021: vii). The nominal form is said to have been coined by Diffloth, as a calque of the French “*impressif*” (Diffloth 2021: vii), with the sense that the nominal form refers to a word class, and is seen as a “technical term” (Diffloth 2021: ix). The titles of two recent volumes focusing on South Asia (*Expressives in the South Asian Linguistic Area*, Badenoch & Choksi eds. 2021; *Expressive morphology in the languages of South Asia*, Williams ed. 2021) suggest that both forms of “expressive” are still in use for descriptions of languages of the larger linguistic area.

### **3.2. Onomatopoeic, onomatopoeia**

The terms “onomatopoeia” (as a substantive) and “onomatopoeic” (and variants, as an adjective) are both found in the descriptions. There are two patterns found: a) the

---

<sup>16</sup> These include *dʰiriri* ‘shouting loudly’, *suk-suk* ‘weeping with voice controlled’, *rinini* ‘insects’ sound’, *sanana* ‘sound of water flowing’.

term is applied to sound-imitative words, conforming to the traditional definition; b) the term is extended beyond sound to other sensory inputs.

As an example of the first situation, van Driem uses the term, as a fully spelled out category label (“onomatopoeia”) in the glossaries appended to his Limbu (van Driem 1987) and Dumi (van Driem 1993) grammars. Because the material is found in the glossary, there is no accompanying text explaining the intent behind its use. The lexemes classified as such do give us a sense of the patterns, such as *pətslək-pətslək* ‘slosh-slosh’ (van Driem 1987: 489), *pyak* ‘slap’ (van Driem 1987: 494), *ya.rrrrr* ‘grrr’ (van Driem 1987: 545), and their glosses show that they are sound imitations.

In the description of Wambule (Opgenort 2004), the term “onomatopoetic” appears in the glossary: lexemes classified as nouns or interjections are occasionally qualified as onomatopoetic, as in the following: “**cikul** *n* onomatopoetic sound associated with the dashing of an egg, slosh” (Opgenort 2004: 580) and “**hui** *interj* onomatopoetic cry” (Opgenort 2004: 618).

Four grammars use “onomatopoetic” (and variants) for more than just sound.

The Bantawa grammar by N.K. Rai (1984), which is the earliest Kiranti grammar to use the term “onomatopoetic”, is one of them. It describes a subclass of complex nouns (“onomatopoetic nouns”, Rai 1984: 64) and adverbs (“onomatopoetic adverbs”, Rai 1984: 153), of which a subtype are “triplicated onomatopoetic adverbs”. Of the onomatopoetic nouns, it is said that “[t]he sound, manner or an action of any object is imitated in this type of nouns” (Rai 1984: 64). The list of examples is mostly made up of natural sounds and the names of certain insects.

In the same grammar, “onomatopoetic adverbs” are described as follows: “Manners of some actions are realized through imitation” (Rai 1984: 153). Some examples are *suiyapni* ‘abruptly’, *bhorakni* ‘very loudly.’<sup>17</sup> An interesting subtype of “onomatopoetic adverbs” are those called “triplicated onomatopoetic adverbs”. Of these, it is said that “the whole triplicated adverb is an imitation of some action, object or feeling, etc.” (Rai 1984: 154). Triplicated onomatopoetic adverbs use the same pattern as triplicated adverbs based on nominal and verbal roots, but are formed from onomatopoeic stems which do not occur independently.

In the body of the grammar of Yamphu (Rutgers 1998), Rutgers devotes no discussion to ideophonic lexemes, apart from two comments about the phonotactics

---

<sup>17</sup> Thulung does not have, to my knowledge, any equivalent ideophones or adverbs, and these cannot therefore be related to a type in § 2.2.

of onomatopoeic words, with *gujiguji* ‘dark, dusk-like’ given as one example (Rutgers 1998: 20). The gloss for this example makes clear that “onomatopoeic” here is used for more than just sound-imitative words. By contrast, we find ‘adv. onom’ in the glossary, where all the examples are for sound.

T.M. Rai’s grammar of Koyee (2015) uses the term “onomatopoeia” for lexemes which, given their glosses, clearly extend beyond the sensory domain of sound: the list provided (Rai 2015: 243–244) includes glosses such as ‘way one smells’, ‘way one licks’, ‘way something shines’, in addition to words that evoke the sense of sound.

There is a strong areal precedent for this extended usage: Emeneau used the term “onomatopoeics” to refer to iconic patterns found across the South Asian linguistic area (Emeneau 1969), and made clear that the sensory scope extends beyond sound:

Semantic identification of the class is more tentative even than formal identification. We are dealing only in the most marginal way with blatantly sound-imitative forms (like English choo-choo or the like). Perhaps it would be more just to say that the class denotes varied types of sensation, the impingement of the material world, outside or within the person, upon the senses-not merely the five conventionally identified senses, but all the feelings, both external and internal. (Emeneau 1969: 284)

His use of the term, associated with patterns, some of which are also found in Kiranti, may explain the term’s endurance for ideophones, despite its traditional limitation to sound.

Masica (1991), in his survey of Indo-Aryan languages, uses the term “onomatopoeia” as well, making clear that he considers it too limited, as these lexemes extend beyond sound “to sensations of many other kinds - visual and tactile” (Masica 1991: 79). Note that in a 1980 re-edition of formerly published essays, including that of 1969, Emeneau prefers the term “expressive” to “onomatopoeic” (of which ideophones are a “subclass in which the symbolism is phonological”), presumably reflecting the fact that “onomatopoeics” seems too limited in scope (Emeneau 1980: 7).

### **3.3. Phonesthetic**

In his grammar of Hayu, Michailovsky (1988) uses the term “phonesthetic words” (in the French text, “mots phonesthétiques”). For the most part these lexemes have an



adverbial use and show total or partial reduplication, such as *gramgram* “(eat) with big mouthfuls” and *krap-krap* “(eat) voraciously”. Michailovsky writes that “phonesthetic’ is [my trans.] “preferable to onomatopoeic for a class of words which are characterized by their use of their phonic nature, which is often but not always unusual, to reinforce their meaning” (1988: 72). These words are “phonologically anomalous” (Michailovsky 1988: 72), with a high frequency of consonant clusters, and pronounced with a bit of emphasis or an interruption in the intonational curve. The section on Hayu phonesthetic words is only a page long, but the description makes an explicit connection with sound symbolism, via the reference to onomatopoeia and the term “phonesthetic” itself, as well as the accompanying explanation of the unusual phonological features which are exploited for the expression and support of the lexemes’ meaning.

This is the only occurrence in the database examined of the term “phonesthetic” as a primary term; as seen above, Weidert & Subba (1985) use it as a paraphrase for “expressive” in their Limbu grammar upon their introduction of the latter.

### 3.4. *Paralexeme*

The term “paralexeme” is first introduced for Kiranti in an article on Bantawa (Rai & Winter 1997), referring to “items whose phonological shape differs from forms in the basic lexicon” (Rai & Winter 1997: 130). In their article, the term is applied to a type of triplicated verbal adjuncts. These are adverbial structures with a nucleus (which they label K) which can be nominal, verbal, or paralexemic, yielding a KKK form followed by an additional “deverbative” suffix *-wa*. The paralexemic nuclei can be recognized by their characteristic phonotactic constraints.

Although the focus of the article is triplicated forms, other words, on the basis of their morphophonological, phonological and phonotactic features, are also considered by the authors to be paralexemic: the authors write of “some twenty configurations characterized by onsets  $C_1C_2a$ - followed by a sequence  $-C_3C_3-$ ; only the continuants *-l-*, *-y-* and *-w-* can fill the position  $-C_2-$ . For  $-C_3-$ , *-y-*, *-p-*, *-t-*, *-k-*, *-m-* and *-ng-* are attested” (Rai & Winter 1997: 131).<sup>18</sup>

---

<sup>18</sup> Although for Thulung, I have resorted to subtyping ideophones on the basis of their morphophonological templates, I do not find any evidence for any particular ideophonic root templates; the paralexemic syllable types presented for Bantawa by Rai & Winter (1997) and by Doornenbal (2009, see following paragraphs) have not to my knowledge been posited for other Kiranti languages.

As far as the scope of the term is concerned, it is used in a way that suggests it refers to both the roots which show special constraints, and also to the output of the triplicated structure. This can be seen in the following citation: “[F]orms may be transferred from one component of the opposition ‘lexemes — paralexemes’ to the other, which makes it possible that a lexemic nucleus may be expanded into a paralexemic triplet” (Rai & Winter 1997: 132), in other words, some elements are paralexemic by virtue of their phonology and phonotactics, but others can become paralexemic when inserted into a distinctive template.

Doornenbal, in his (2009) grammar of the same language, reprises the term “paralexeme”. He proposes two syllable types –  $C_oC_aV$  and  $C_oC_aVC_f$  (for which  $C_o$  is an onset consonant,  $C_a$ , an approximant, and  $C_f$ , a final consonant) – which can be used to classify (non-borrowed) words as “paralexemes that often have an ideophonic or onomatopoeic aspect to their meaning” (2009: 43).

While “paralexeme” is the dominant term used to refer to these Bantawa elements, Doornenbal (2009) uses a number of other terms, in adjectival form, in connection with the same lexemes. This is seen in the citation in the previous paragraph, where “ideophonic” and “onomatopoeic” are used to complete the interpretation of “paralexeme”. We find the following terms in the grammar: “ideophonic” (collocating with “association”, “usage”, “aspect”, “value”); “onomatopoeic” (collocating with “association”, “aspect”, “qualities”); “expressive” (collocating with “adverbs”, “qualities”). We additionally find references to “mimetics”: “The term ‘paralexemes’ corresponds to ‘mimetics’” (Doornenbal 2009: 36); “The paralexemic class of words typologically are on par with mimetics, as described by Ito & Mester (1995)” (Doornenbal 2009: 303).

As can be seen from the various citations provided, these terms are often found in the same sentences, serving as paraphrases of each other (cf. “expressive adverbs that are onomatopoeic or mimetic” (Doornenbal 2009: 298)). The coupling of terms has a bridging effect, associating lexemes which are identified as having a certain number of significant and marked characteristics with a wide range of terminology, ensuring that this will trigger recognition regardless of readers’ terminological backgrounds and stances.

The use of the term “paralexeme” does not occur elsewhere in my corpus of Kiranti materials, not even in an article on Chintang co-authored by N.K. Rai (Rai et al. 2005), author of the Rai & Winter (1997) study. The term essentially captures the fact that certain words (and roots) have characteristics which do not concord with the prosaic lexicon – in the case of Bantawa, it is made clear that this is on account of a syllable type which confers an iconic value to words which share it.

### 3.5. Ideophone

The first occurrence of ideophone in the corpus used here is in the form of a gloss in examples in Bickel (1996) on Belhare. An example is reproduced in what follows: in (8), a lexical gloss follows the abbreviation “IDEOPH”, but sometimes “IDEOPH” occurs alone.

- (8) Belhare (Kiranti; Bickel 1996: 222)  
*u-niūa*      *u-phokg-ep-phu*      *bhutbhuti-bu*      *lis-e.* < KP28b >  
3POSS-mind    3POSS-belly-LOC-REP    IDEOPH(strange)-REP    be-PT  
'He felt strange in (both) mind and stomach.'

The abbreviation in examples (and in the abbreviations list) is the only use of “ideophone” in the work, and it is not accompanied by any further description.

Ideophone in the form of a gloss abbreviation is also found in work by Ebert (Ebert 1997a, 1997b, 2000) on Athpare and Camling. It features in the examples illustrating the two grammatical descriptions, as well as in the Camling texts. Interestingly, ideophones are not included in the glossaries provided for both languages, suggesting Ebert considers them to have a different status from the prosaic lexicon which makes up the glossaries.

For Kiranti descriptions, “ideophone” first appears as a full term in an article by Rai et al. (2005), on Chintang triplicated ideophones. “Ideophone” here is firmly anchored in the typological literature, accompanied by definitions by Doke (1935) and Voeltz & Kilian-Hatz (2001). For Chintang, the term refers to one possible type of root (the others are nouns and verbs) which is the input of triplication, and the output is treated as a member of the class of adverbs. It is interesting, from a terminological point of view, that when the term is first used in the introduction of the article, it is as “ideophonic (onomatopoetic) roots” (Rai et al. 2005: 205). This parenthetical addition is in line with the novelty of the term in descriptions of Kiranti languages and the need to provide context for the concept.

Following the publication of the Rai et al. (2005) article, the term “ideophone” (or “ideophonic”) is found as the main term for Puma (Sharma 2014), Yakkha (Schackow 2015), Chintang (Paudyal 2015), Khaling (Lahaussais 2017a) and Thulung (Lahaussais 2023).

In many of the descriptions which use the term “ideophone”, it is made clear that the ideophonic element is the base, often as an alternative to nominal and verbal bases, which is found in a specific pattern (the descriptions frequently focus on the marked triplicated pattern). The descriptions which make an explicit claim along these lines, using the term “ideophone” or “ideophonic”, are those of Chintang (Rai et al. 2005; Paudyal 2015), Bantawa (2009) and Yakkha (Schackow 2015). Other descriptions using the term do not make any explicit claims about which element is ideophonic, although the grammar of Puma lists ideophone as a type of “grammatical category”, a term which Sharma opposes to the lexical word classes of nouns, adjectives, adverbs, verbs, pronouns and numerals (Sharma 2014: 100).

Noteworthy are two articles by Caughley on ideophones in Chepang, a related language of Nepal outside the Kiranti subgroup: the first (Caughley 1997) uses “ideophone” in the article, which compares the characteristics of vowel gradation in Chepang with data from Sunwar (a Kiranti language); the second (Caughley 2002) focuses on the ideophones of Chepang. Caughley writes of the previous use of the term onomatopoeia for these lexemes that “[h]owever an examination of the data shows that, while many roots in this subclass are indeed sound-imitative, a considerable number have no reference at all to sounds” (Caughley 1997: 96). He continues a little further with “I will henceforth use the term ‘ideophone’ for this special subset of adverbs which includes both sound-imitative and non-sound imitative roots.” While neither article by Caughley is cited with any frequency<sup>19</sup> in the materials in my database, it may have had an influence on linguists working in Nepal.

### 3.6. *Adverb*

In some grammars, ideophones are placed in the class of adverbs,<sup>20</sup> where they may or may not be further sorted into subclasses.

---

<sup>19</sup> The exceptions are Doornenbal (2009) and Schackow (2015), both of whom cite Caughley’s 1997 article.

<sup>20</sup> There is a long precedent for this: ‘adverb’ has been used as a catch-all term throughout the history of linguistic description for misfit classes (see e.g. Odoul 2019) such as interjections (Lahaussais 2019a).

Allen's (1975) Thulung grammar has a large category of "expressive resources", within which the different subclasses of ideophones are labeled as different types of "adverb".

Opgenort's (2004) Wambule grammar contains ideophones of several patterns, one subtype of which is labeled "adverbial proclitics of manner" (corresponding to the Thulung ideophonic preverbs in § 2.2. above). His grammar provides a list of these (Opgenort 2004: 243), but additional lexemes fitting the same pattern (both in terms of morphophonology and semantics) are found in the glossary, where they are glossed simply as "adverb". Other lexemes which map onto other patterns receive the gloss *adv*, such as the partially triplicated<sup>21</sup> and reduplicated forms, which are glossed *adv cmp* and *adj cmp* ("compounded adverb" and "adjective", respectively). Opgenort thus uses the term "adverb" (either alone, as "compounded adverb", or as "adverbial proclitics") for all the lexemes that map onto ideophonic patterns in other languages.

Tambahang's grammar of Chhatthare Limbu labels "manner adverbs" lexemes which are "formed by reduplication of the first syllable" (2017: 45).<sup>22</sup> Both the form, which matches the reduplicated pattern found throughout Kiranti, and the sensory domains in the glosses suggest that these are ideophonic lexemes.

In other grammars, the main term "adverb" is modified by a term which points to ideophonic properties. This is the case of "expressive adverbs" (Weidert & Subba 1985; Rai 2016) and "onomatopoetic adverbs" (Rai 1984; Rutgers 1998).

### 3.7. Evolution of terminology

The terms presented in the sections above can be summed up as follows: The earliest term we find in the database of descriptions is "expressive", as an adjective modifying "resources" (Allen 1975); the term is used ten years later as both an adjective and a substantive (Weidert & Subba 1985). "Onomatopoetic" first comes up in N.K. Rai (1984), in adjectival form; in substantival form, it is first found in work by van Driem (1987, 1993). Two other authors in the database use it after this as their primary term: Rutgers (1998) and T.M. Rai (2016). Michailovsky's "phonesthetic words" (1988), which first appears as an alternative to "expressive" in Weidert & Subba

---

<sup>21</sup> An example is "brwakwakwak [attested with kakcam] *adv* manner of breaking open like popcorn" (Opgenort 2004: 560).

<sup>22</sup> Examples are *yanyan* 'lightly', *təktək* 'straight', *pəkpek* 'disorderly manner'.

(1985), does not get recirculated as a primary term in this database. “Paralexeme” is introduced by Rai & Winter (1997), taken up by Doornenbal (2009) for the same language, Bantawa, and is linked to a syllable type, which can appear as the root in certain marked patterns. Finally, “ideophone” appears as a gloss (Bickel 1996) then as full term (Rai et al. 2005), and becomes the most common term thereafter.

An interesting find in looking through the Kiranti materials is that most of the grammarians accompany the main term for ideophonic lexemes with other related terms, something which does not occur with other word classes. The use of the primary term is followed by alternative terms, paraphrases, and sometimes a negative comparison. The authors who use such a comparative or paraphrasing strategy are listed below, along with the terms:

- Allen (1975) makes an explicit connection between the “expressive resources” of Thulung and ideophones;
- Weidert & Subba (1985) provide clarification upon first using the term “expressive” by segueing with the term “phonaesthetic” (“expressive or phonaesthetic adverbs”, Weidert & Subba 1985: 15);
- Michailovsky (1988) states that his chosen term “phonesthetic words” is a preferable term to “onomatopoeic”, which is inadequate;
- N.K. Rai & Winter (1997) mention the iconicity inherent in “paralexemes”, and refer to onomatopoeic as a term which “does not do much good: what is it that is imitated by these forms?” (Rai & Winter 1997: 132);
- N.K. Rai et al. (2005) add onomatopoeic in parentheses, as a clarification, upon the first use of ideophonic in their article’s introduction;
- Doornenbal (2009) uses a wide range of terminology to clarify the term “paralexemic” and the words it describes: “ideophonic”, “onomatopoeic”, “expressive”, “mimetic”, creating a semantic web linking these terms;
- Schackow (2015) clarifies some uses of “ideophonic” but specifying that they involve an “ideophonic component (i.e., an iconic relationship between the concept expressed and the phonological form)” (Schackow 2015: 179).

The presence of such paraphrases and comparisons suggests that individual terms are not felt to be sufficiently well-defined or -described in order to convey the markedness and unique characteristics of these words. The use of multiple equivalent terms helps fine-tune the description, triangulate the properties of these words, and ensure that

the reader can map the features onto alternative terminology which may be more familiar. The overall sense is that a single term is not enough to convey the whole picture, something likely to change with the apparatus that now surrounds the term “ideophone” and which includes not only solid definitions (Dingemanse 2019; Dingemanse 2023) but also the heuristic tool of the implicational hierarchy (Dingemanse 2012; McLean 2021).

#### 4. Discussion

Section 3 has provided discussion for the use of terms for ideophones. As we saw, apart from synonyms and comparanda, which are useful in clarifying what authors of a description mean by the use of a specialized term, there is very little material in the descriptions that explicitly discusses the terminological choices. I attempt here to map out some of the advantages and disadvantages of the most frequent terms found in our corpus. Because of their low frequency, used for a single language and not having been adopted by subsequent linguists working on different languages of the subgroup, I do not discuss “phonesthetic words” or “paralexemes”.

As far as “expressive” is concerned, there is a contrast between adjectival and nominal uses. Diffloth (1976: 263–44, fn 2) offers the following definition of “expressives” (in the nominal form),<sup>23</sup> in contrast with onomatopoeia and ideophones:

[O]nomatopoetic forms are those displaying acoustic symbolism and having syntactic and morphological properties totally different from those of verbs and nouns. *Ideophones* are words displaying phonological symbolism of any kind (acoustic, articulatory, structural) and having distinct morphosyntactic properties; ideophones include onomatopoetic forms as a subclass. *Expressives* have the same morphosyntactic properties as ideophones, but their symbolism, if such exists, is not necessarily phonological; expressives contain ideophones as a subclass.

It is not clear from the contrastive definition how far the class of “expressive” is meant to extend, and whether the elaborate expressions and psycho-collocations (Matisoff 1986) of Mainland Southeast Asia are included, as they seem to be in some of the contributions to a volume edited by Williams (2013). Even though Diffloth’s term is

---

<sup>23</sup> This can be contrasted with Jakobson’s description of *expressive function*, which “aims a direct expression of the speaker’s attitude toward what he is speaking about. It tends to produce an impression of a certain emotion whether true or feigned” (Jakobson 1960: 354).

intended to point to a word class, it is one with fairly loose boundaries, able to accommodate a range of material.

“Onomatopoeia” as a term has the advantage of being linked with traditional (and school) grammar, where it signals a non-arbitrary connection between the sound of a word and its meaning.<sup>24</sup> This is the use which is applied in some of the grammars we have seen. The problem is that it is a category that can encompass words of different syntactic functions, so that nominal forms (like onomatopoeic names of animals, as in N.K. Rai’s (1984) Bantawa grammar) and adverbial forms (in many cases, sound ideophones) are given the same label. An additional problem is that some grammars extend the term “onomatopoeia” to non-sound ideophones, and as a result the term is used with different interpretations, even within the same subgroup.

When “adverb” is used in grammars in the corpus to refer to ideophones, they constitute one or more subclasses with specific properties. The subclasses’ properties are generally conveyed by the modifier that accompanies the term, which can focus on various features: the form of the word (“reduplicated”, “with -maksī”, etc.), on their semantic properties (“expressive adverbs”), or other attributes. Unlike languages for which ideophones form a clear word class, as is the case in many African languages (see the contributions by Guérois, Meyer, Quint and Treis, this volume), Kiranti ideophones are probably at best a word class made up of a number of rather different subclasses. Using the term “adverb” (+ modifier) is a convenient way to focus first on their shared morphosyntactic properties,<sup>25</sup> and then to specify what characteristics define the various subclasses, through the use of the modifier. The disadvantage of labeling ideophones as “adverbs” is that they can get lost in a fairly large super-category and not be given their proper due, in terms of their special morphophonological patterns, semantic properties, and the like.

An advantage for more recent grammars which use “ideophone” is that the term inscribes those descriptions directly into a cross-linguistic landscape of research into similar phenomena. With an ever-increasing number of descriptions of ideophones around the world, being able to participate in the wider discussion is also tied to the choice of terminology, for the simple reason that using a widely-accepted term makes one’s work easier to identify and relate to. Choosing a term such as “ideophone” suggests an acceptance of pre-existing tools for carrying out ideophone research: this can be engaging with implicational hierarchies for sensory domains targeted by ideophones (Dingemans 2012; McLean 2021), taking as a starting point existing definitions (Doke 1935; Voeltz & Kilian-Hatz 2001; Dingemans 2019), and deriving

---

<sup>24</sup> See Moore (2015) for a history of use of the term in English prior to the 20th century.

<sup>25</sup> See Dingemans (2018: 2) on the mapping between characteristics of ideophones and the terms applied to them.



ideas for what interesting behaviors to report on from what is described in other languages. An example of the latter is the patterns of consonant and vowel gradation in ideophones in some languages (see Marsault, this issue, for the same phenomenon in Omaha), which is found within the domain of Trans-Himalayan languages of Nepal (Caughley 1997), but not described for Kiranti. Although adopting a comparative term may carry the risk of superficially overlooking specificities of patterns in a specific language, it also carries significant advantages.

## 5. Conclusion

This article has provided a documentation of the diverse terminology used for ideophonic lexemes in Kiranti language descriptions. The starting point for the investigation was a number of morphophonological patterns shown by lexemes, generally but not always with an adverbial function, with ideophonic properties in Kiranti languages (Lahaussais 2023), which were used to identify similar material in grammars and to collect the associated terminology.

This study revealed a wide range of terminology, in some cases accompanied by alternative terms and paraphrases in order to clarify a term which might seem opaque or marginal. Apart from these synonyms and comparanda, terms were rarely accompanied by explanations of what was meant by the term.

The introduction to this article mentioned that one of the difficulties in comparing ideophones in Kiranti languages is due to the profusion of terminology and difficulty in identifying materials to compare. Another difficulty was methodological: grammars generally try to provide appended glossaries which are relatively complete, as far as nouns, verbs and other major parts of speech are concerned (this is aided by the use of word lists of the non-Swadesh variety),<sup>26</sup> but not for ideophones. Because of the translational difficulties associated with ideophones (e.g. Msimang & Poulos 2001: 235), it would be difficult to devise a list which could be used for their elicitation. Furthermore, even if such a word list could be produced for elicitation, the glosses would need to be precise enough to allow comparison, which is a tall order given the nature of ideophones.<sup>27</sup> A corollary of this appears to be that when ideophonic

---

<sup>26</sup> An example is the Living Tongues *Language Sustainability Toolkit* (Daigneault et al. 2022); others, with regional orientations, can be found in the TULQuest archive (Lahaussais 2019b).

<sup>27</sup> The glosses given for some Koyee ‘onomatopoeia’ are impossible to use for any study of possible cognacy, as they are semantically quite vague: see examples in § 3.2. above, as well as glosses such as

lexemes are provided (the same goes for interjections), it is in the form of a sampling and presumably far from complete. This suggests a real need for better methodological tools for the collection of Kiranti ideophones: one potential venue is to create a list of ideophonic materials assembled from all existing descriptions of Kiranti languages, which can be used to trigger associations in consultants on the basis of either phonology or semantics.<sup>28</sup> The collection of new materials must of course be coupled with good examples, which provide as many contexts for use as possible.

An ulterior motive for the study presented here, and for that in Lahaussais (2023), is to determine whether there is any shared matter, beyond the attested shared patterns, for Kiranti ideophonic lexemes, and if so (there is distinct evidence of this already), whether the shared matter is a result of cognacy or borrowing. The examination of the terms used and the types of patterns they are associated with has generated a larger collection of Kiranti ideophones that I previously had access to, and has set the groundwork for a study of their actual forms with a view to better understanding their diachrony.

## Acknowledgments

I thank two anonymous reviewers as well as the co-editors of this special issue for their insightful comments on this contribution. I would also like to thank organizers and audience members of the combined LSN-43/SALA-36 conferences, held in Kathmandu, in November, 2022, where I presented some of this material. This work is supported by a public grant overseen by the IdEx Université Paris Cité (ANR-18-IDEX-0001) as part of the Labex Empirical Foundations of Linguistics - EFL.

---

‘manner of something falling’, ‘manner someone beats something’, ‘when everything is finished/disposed of’ (Rai 2015: 243–244).

<sup>28</sup> I have experimented with this technique successfully in increasing my list of Thulung ideophones in December 2022: a spreadsheet of ideophones, sorted by subclass, from 9 different languages was used as a prompt and allowed me to collect many new ideophones, either because they triggered a sound connection, or because the semantics elicited a Thulung form.

## Abbreviations

2 = 2 <sup>nd</sup> person	IDEOPH = ideophone	POSS = possessive
3 = 3 <sup>rd</sup> person	IDEO.PRVB = ideophonic preverb	PST = past
ABL = ablative	IMP = imperative	PT = past
DER = derivational marker	INF = infinitive	REP = report marker
DU = dual	INT = intensifier	SG = singular
HS = hearsay marker	LOC = locative	TEMP = temporal sequencer
IDEO = ideophone	NEG = negative	

## References

- Abbi, Anvita. 2018. Echo formations and expressives in South Asian languages. A probe into significant areal phenomena. In Aina Urdze (ed.), *Non-prototypical reduplication*, 1–34. Berlin / Boston: De Gruyter Mouton.
- Allen, Nicholas J. 1975. *Sketch of Thulung grammar*. Ithaca, NY: China-Japan Program, Cornell University.
- Bickel, Balthasar. 1996. *Aspect, mood, and time in Belhare*. Zurich: University of Zurich.
- Borchers, Dörte. 2008. *A grammar of Sunwar: descriptive grammar, paradigms, texts and glossary*. Leiden: Brill.
- Caughley, Ross. 1997. Semantically related vowel gradation in Sunwar and Chepang. *Papers in Southeast Asian Linguistics* 14. 95–101.
- Caughley, Ross. 2002. Ideophones in Chepang: Their nature and subcategorization. *Gipán: Tribhuvan University Papers in Linguistics* 2 (May). 16–24.
- Chelliah, Shobhana & Mary Burke & Marty Heaton. 2021. Using interlinear gloss texts to facilitate cross-language comparison and improve language description. *Indian Linguistics* 82(1–2). 1–24.
- Dahal, Ballabh Mani. 1999. Phonesthetic elements in Nepali. In Yogendra P. Yadava & Warren W. Glover (eds.), *Topics in Nepalese linguistics*, 90–102. Kathmandu: Royal Nepal Academy.
- Daigneault, Anna Luisa & D. Bögre Udell & K. Tcherneshoff & Gregory D. S. Anderson. 2022. The language sustainability toolkit. Living Tongues Institute for Endangered Languages and Wikitongues. <https://livingtongues.org/language-sustainability-toolkit/>. (Accessed 2023.01.19).
- Diffloth, Gérard. 2021. Foreword. In Nathan Badenoch & Nishant Choksi (eds.), *Expressives in the South Asian linguistic area*, vii–x. Leiden: Brill.

- Diffloth, Gérard. 1976. Expressives in Semai. In *Austroasiatic Studies Part I*, vol. 13, 249–264. Honolulu: University of Hawai'i Press.
- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6(10). 654–672.
- Dingemanse, Mark. 2018. Redrawing the margins of language: Lessons from research on ideophones. *Glossa: a journal of general linguistics* 3(1). 1–30. <https://doi.org/10.5334/gjgl.444>.
- Dingemanse, Mark. 2019. “Ideophone” as a comparative concept. In Kimi Akita & Prashant Pardeshi (eds.), *Ideophones, mimetics and expressives*, 13–34. Amsterdam: John Benjamins.
- Dingemanse, Mark. 2023. Ideophones. In Eva van Lier (ed.), *Oxford handbook of word classes*. Oxford: Oxford University Press, 466–476.
- Doke, Clement Martyn. 1935. *Bantu linguistic terminology*. London: Longmans.
- Doornenbal, Marius. 2009. *A grammar of Bantawa: Grammar, paradigm tables, glossary and texts of a Rai language of Eastern Nepal*. Utrecht: LOT.
- Ebert, Karen. 1994. *The structure of Kiranti languages*. Zurich: University of Zurich.
- Ebert, Karen. 1997a. *Camling (Chamling)*. Munich: Lincom Europa.
- Ebert, Karen. 1997b. *A grammar of Athpare*. Munich: Lincom Europa.
- Ebert, Karen. 2000. *Camling texts and glossary*. Munich: Lincom Europa.
- Emeneau, Murray. 1969. Onomatopoeics in the Indian Linguistic Area. *Language* 45(2.1). 274–299.
- Emeneau, Murray. 1980. *Language and linguistic area: essays*. Stanford, CA.: Stanford University Press.
- Gerber, Pascal & Selin Grollman. 2018. What is Kiranti? *Bulletin of Chinese Linguistics* 11. 99–152.
- Grierson, George A. (ed.). 1909. *Linguistic survey of India*. Vol. III.1: *Tibeto-Burman family: General introduction, specimen of the Tibetan dialects, the Himalayan dialects, and the North Assam group*. Delhi: Motilal Banarsidass.
- Hodgson, Brian H. 1857a. Comparative vocabulary of the languages of the broken tribes of Nepal [Vocabulary and grammar of the Vayu tribe]. *Journal of the Asiatic Society of Bengal* 26. 317–522 [372–485].
- Hodgson, Brian H. 1857b. Comparative vocabulary of the languages of the broken tribes of Nepal [Vocabulary of the Bahing tribe]. *Journal of the Asiatic Society of Bengal* 26. 317–522 [486–522].
- Hodgson, Brian H. 1858. Comparative vocabulary of the languages of the broken

- tribes of Nepal (continued from vol 26) [Grammar of the Bahing tribe]. *Journal of the Asiatic Society of Bengal* 27. 393–456 [393–442].
- Jakobson, Roman. 1960. Closing statement: Linguistics and poetics. In Thomas Sebeok (ed.), *Style in language*, 350–377. Cambridge, MA: The Technology Press of Massachusetts Institute of Technology.
- Kelly, Barbara & Aimée Lahaussais. 2021. Chains of influence in Himalayan grammars: models and interrelations shaping descriptions of Tibeto-Burman languages of Nepal. *Linguistics* 59(1). 207–245. <https://doi.org/10.1515/ling-2020-0255>.
- Lahaussais, Aimée. 2003. Nominalization and its various uses in Thulung Rai. *Linguistics of the Tibeto-Burman Area* 26(1). 33–57.
- Lahaussais, Aimée. 2017a. Ideophones in Khaling Rai. *Linguistics of the Tibeto-Burman Area* 40(2). 179–201. <https://doi.org/10.1075/ltba.17005.lah>.
- Lahaussais, Aimée. 2017b. Relativization strategies and alignment in Thulung Rai. *Cahiers de Linguistique Asie Orientale* 46(1). 73–100. <https://doi.org/10.1163/19606028-04601003>.
- Lahaussais, Aimée. 2019a. Interjection. In Bernard Colombat & Aimée Lahaussais (eds.), *Histoire des parties du discours*, 455–473. Leuven: Peeters.
- Lahaussais, Aimée. 2019b. The TULQuest linguistic questionnaire archive. In Aimée Lahaussais & Marine Vuillermet (eds.), *Methodological tools for linguistic description and typology*, 31–44. Honolulu: University of Hawai'i Press.
- Lahaussais, Aimée. 2021. Glossing in the Linguistic Survey of India: Some insights into early 20th century practices. *Historiographia Linguistica* 48(1). 25–59.
- Lahaussais, Aimée. 2023. Ideophonic patterns in Kiranti languages and beyond. *Folia Linguistica*.
- Lehmann, Christian & Elena Maslova. 2004. Grammaticography. In Geert Booij, Christian Lehmann, Joachim Mugdan & Stavros Skopeteas (eds.), *Morphologie: Ein Handbuch zur Flexion und Wortbildung*, 1857–1882. Berlin / New York: W. de Gruyter.
- Majeed, Javed. 2019a. *Colonialism and knowledge in Grierson's Linguistic Survey of India*. London / New York: Routledge.
- Majeed, Javed. 2019b. *Nation and region in Grierson's Linguistic Survey of India*. London / New York: Routledge.
- Masica, Colin. 1976. *Defining a linguistic area: South Asia*. Chicago: University of

- Chicago Press.
- Masica, Colin. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Matisoff, James. 1986. Hearts and minds in South-East Asian languages and English: An essay in the comparative lexical semantics of psycho-collocations. *Cahiers de Linguistique Asie Orientale* 15(1). 5–57.
- McLean, Bonnie. 2021. Revising an implicational hierarchy for the meanings of ideophones, with special reference to Japonic. *Linguistic Typology* 25(3). 507–549.
- Michailovsky, Boyd. 1988. *La langue hayu*. Paris: Editions du CNRS.
- Michailovsky, Boyd. 2017. Kiranti languages. In Graham Thurgood & Randy LaPolla (eds.), *The Sino-Tibetan languages*, 646–679. 2nd edn. Oxford / New York: Routledge.
- Moore, Colette. 2015. An ideological history of the English term onomatopoeia. In Michael Adams, Laurel J. Brinton & R. D. Fulk (eds.), *Studies in the history of the English language. VI: Evidence and method in histories of English*, 307–321. Berlin/Munich/Boston: De Gruyter Mouton.
- Msimang, Christian T. & George Poulos. 2001. The ideophone in Zulu: A re-examination of conceptual and descriptive notions. In F. K. Erhard Voeltz & Christa Kilian-Hatz (eds.), *Ideophones*, 235–249. Amsterdam / Philadelphia: Benjamins.
- Odoul, Marie. 2019. Adverbe. In Bernard Colombat & Aimée Lahaussais (eds.), *Histoire des parties du discours*, 365–386. Leuven: Peeters.
- Opgenort, Jean Robert. 2004. *A grammar of Wambule: Grammar, lexicon, texts and cultural survey of a Kiranti Tribe of Eastern Nepal*. Leiden: Brill.
- Opgenort, Jean Robert. 2005. *A grammar of Jero, with a historical comparative study of the Kiranti languages*. Leiden / Boston: Brill.
- Paudyal, Netra Prasad. 2015. *Aspects of Chintang syntax*. University of Zurich.
- Pokharel, Madav P. 1993. The grammar of onomatopoeia in Nepali. *Nepalese Linguistics* 10. 10–34.
- Rai, Netra Mani. 2016. *A grammar of Dumi*. Kathmandu, Nepal: Tribhuvan University (PhD dissertation).
- Rai, Novel Kishore. 1984. *A descriptive study of Bantawa*. Poona: Deccan College.
- Rai, Novel Kishore, Balthasar Bickel, Martin Gaenszle, Elena Lieven, Netra Prasad Paudyal, Ichchha Purna Rai, Manoj Rai & Sabine Stoll. 2005. Triplication and ideophones in Chintang. In Yogendra P. Yadava (ed.), *Current issues in Nepalese linguistics*, 205–209. Kirtipur: Linguistic Society of Nepal.

- Rai, Novel Kishore & Werner Winter. 1997. Triplicated verbal adjuncts in Bantawa. In David Bradley (ed.), *Tibeto-Burman languages of the Himalayas*, 119–134. Canberra: Pacific Linguistics.
- Rai, Tara Mani. 2015. *A grammar of Koyee*. Tribhuvan University.
- Rai, Vishnu Singh. 2012. *A grammar of Chamling*. Bern: University of Bern.
- Rutgers, Roland. 1998. *Yamphu: Grammar, texts, lexicon*. Leiden: Research School CNWS.
- Schackow, Diana. 2015. *A grammar of Yakkha*. Berlin: Language Science Press.
- Schlemmer, Grégoire. 2021. Rai: Following the ancestors and managing the otherness. In Marine Carrin (ed.), *Encyclopedia of the religions of indigenous people of South Asia*, 785–792. Leiden / Boston: Brill.
- Sharma, Narayan Prasad. 2014. *Morphosyntax of Puma, a Tibeto-Burman language of Nepal*. London: SOAS, University of London.
- Thurgood, Graham & Randy LaPolla (eds.). 2017. *The Sino-Tibetan languages*. 2nd edn. Oxon: Routledge.
- Toba, Sueyoshi. 1984. *Khaling*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Tolsma, Gerard. 2006. *A grammar of Kulung*. Leiden: Brill.
- Tumbahang, Govinda Bahadur. 2007. *A descriptive grammar of Chhatthare Limbu*. Kathmandu, Nepal: Tribhuvan University.
- Tumbahang, Govinda Bahadur. 2017. *A grammar of Chhatthare Limbu*. Munich: Lincom.
- van Driem, George. 1987. *A grammar of Limbu*. Berlin / New York, Amsterdam: Mouton de Gruyter.
- van Driem, George. 1993. *A grammar of Dumi*. Berlin / New York: Mouton de Gruyter.
- van Driem, George. . 2018. Linguistic history and historical linguistics. *Linguistics of the Tibeto-Burman Area* 41(1). 106–127. <https://doi.org/10.1075/ltba.18005.dri>.
- Voeltz, F. K. Erhard & Christa Kilian-Hatz (eds.). 2001. *Ideophones*. Amsterdam / Philadelphia: Benjamins.
- Weidert, A. & B. Subba. 1985. *Concise Limbu grammar and dictionary*. Amsterdam: Lobster Publications.
- Williams, Jeffrey P. (ed.). 2013. *The aesthetics of grammar. Sound and meaning in the languages of Mainland Southeast Asia*. Cambridge: Cambridge University Press.
- Wolfenden, Stuart. 1933a. A specimen of the Sāngpāng Dialect. *Acta Orientalia* 12.

71–79.

Wolfenden, Stuart. 1933b. Specimen of a Khambu Dialect from Dilpa, Nepāl. *The Journal of the Royal Asiatic Society of Great Britain and Ireland* 4. 845–856.

Wolfenden, Stuart. 1934. A specimen of the Kûlung Dialect. *Acta Orientalia* 13. 35–43.

Wolfenden, Stuart. 1935. A specimen of the Thûlung Dialect. *Journal of the Royal Asiatic Society* 67(4). 629–653.

#### CONTACT

aimee.lahaussais@cnrs.fr



# The sound roots of Umó<sup>n</sup>ho<sup>n</sup>

JULIE MARSAULT

CNRS, UMR 7597 HISTOIRE DES THÉORIES LINGUISTIQUES

Submitted: 31/01/2023 Revised version: 17/01/2024

Accepted: 22/01/2024 Published: 12/06/2024



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article)

## Abstract

This paper presents a corpus-based study of lexemes denoting sounds in Umó<sup>n</sup>ho<sup>n</sup> (oma), a Siouan language of North America. I take as a starting point a list of sound-denoting verbal roots (in short: “sound roots”), presented as onomatopoeia in a paper by Dorsey in 1892, that form a coherent set based on their semantic features – they denote sounds. I describe their morphological and syntactic features and their form-meaning mappings in order to assess (1) whether these features distinguish them from other verbal roots, and (2) how well they fit the cross-linguistic definition of ideophones proposed by Dingemanse in 2019. I show that several salient morphological and syntactic features are repeatedly attested with sound roots. However, the currently available corpus does not provide evidence that the sound roots form a homogeneous class on the morphological and syntactic level, due to the disparity of features attested from one root to the other. Hence I conclude that these roots cannot be considered ideophones in Dingemanse’s sense. Nonetheless, similarities between the sound roots of Umó<sup>n</sup>ho<sup>n</sup> and ideophones in other languages can be observed. They can be grammatically integrated, by contrast with onomatopoeia, and their meaning extends from sound to other sensory domains.

**Keywords:** ideophones; onomatopoeia; Siouan languages; instrumental affixes; consonant gradation; depiction

## 1. Introduction

### 1.1 Background

Native languages of North America have heretofore received little attention in studies on ideophones and are not generally described as having such a lexical class (but see Munro 1998; Duncan 2022; de Reuse 2022). Siouan languages, however, have a notable number of lexemes denoting sounds, which is the first category given in implicational hierarchies of the semantic categories covered by ideophones cross-linguistically (Dingemanse 2012; McLean 2021).

This paper deals with sound-denoting roots in Umó<sup>n</sup>ho<sup>n</sup>(-Pá<sup>n</sup>ka),<sup>1</sup> a Siouan language of the Mississippi Valley branch. It identifies and describes a set of SOUND ROOTS and discusses how well they fit the comparative concept of ideophones as formulated by Dingemanse (2019: 16): “member[s] of an open lexical class of marked words that depict sensory imagery”. In this paper, I use the term “lexical class” in reference to a distinctive class of lexemes that form a coherent group and a subcategory of a part of speech.<sup>2</sup>

Umó<sup>n</sup>ho<sup>n</sup> (Omaha; oma), traditionally spoken by the eponymous tribe, is a critically endangered (almost dormant) language of the Siouan language family, and whose community is located in the current state of Nebraska, United States. It was extensively documented by the Reverend James Owen Dorsey at the end of the 19<sup>th</sup> century, and by linguists and community-based language and culture centers from the 1970s on. There are currently no Umó<sup>n</sup>ho<sup>n</sup>-dominant speakers, but many elderly members of the Umó<sup>n</sup>ho<sup>n</sup> Nation retain various degrees of language knowledge from their childhood and youth. A few diachronic changes can be observed between 19<sup>th</sup>-century Umó<sup>n</sup>ho<sup>n</sup> and modern Umó<sup>n</sup>ho<sup>n</sup> (cf. Marsault 2021: 150-153), none of which are relevant to the present discussion, to the best of my knowledge. (Note, however, that the total amount of available data is biased towards 19<sup>th</sup>-century Umó<sup>n</sup>ho<sup>n</sup>.)

In the remainder of this introduction, I will present the main typological features of Umó<sup>n</sup>ho<sup>n</sup> (Section 1.2) and comment on the type of sources used for my study

---

<sup>1</sup> The Umó<sup>n</sup>ho<sup>n</sup> and the Pá<sup>n</sup>ka are distinct but related tribes, with the same heritage language. The differences are mainly lexical, but each tribe generally refers to its own language as Umó<sup>n</sup>ho<sup>n</sup> or Pá<sup>n</sup>ka. This work is based on old documentation of both varieties and modern documentation of Umó<sup>n</sup>ho<sup>n</sup>.

<sup>2</sup> Dingemanse (2019: 15-16) apparently has a broader understanding of a “lexical class” as “a distinctive stratum of vocabulary”, that can cover several parts of speech.

(Section 1.3). In Section 2, I present my data set of sound roots and how I assembled it. In Section 3, I present morphological and syntactic features of the sound roots, and in Section 4 form-meaning mappings. Based on this description, in Section 5, I compare the sound roots in Umó<sup>n</sup>ho<sup>n</sup> to ideophones in other languages and to the comparative concept. I conclude in Section 6.

## 1.2 Main typological features of Umó<sup>n</sup>ho<sup>n</sup>

Umó<sup>n</sup>ho<sup>n</sup> is a prefixing and head-marking language with a head-final constituent order: the basic word order is verb-final and dependent clauses precede main clauses, as in (1). It has a very rich verbal morphology that reflects areal features of Native North American languages (see Mithun 2015, 2017). In particular, it has applicative prefixes and a series of instrumental prefixes (see Section 3.6) and, despite encoding up to two arguments on the verb, it has no marker for 3<sup>rd</sup> persons.<sup>3</sup> Person marking follows a split intransitive alignment. Verbs constitute the main part of speech in terms of size, since property words are intransitive stative verbs (adjectival verbs). Many verbs are formed by the combination of a bound root with one or several derivational morphemes, especially the instrumental prefixes that I will introduce in Section 3.1. A few post-verbal morphemes trigger an Ablaut of the final vowel /e/ into /a/. For instance, the bound root *tíde* ‘(there is) a drumming sound’<sup>4</sup> becomes *tída* before the proximate/plural enclitic =*í* in (2).

(1) (Dorsey 1890: 71.13 / speaker: Hupetho<sup>n</sup>)<sup>5</sup>

<i>Hó<sup>n</sup>egó<sup>n</sup>ch<sup>h</sup>e</i>	<i>ki</i>	<i>wahó<sup>n</sup> = biamá.</i>
morning	when	move = PP.REPORT
<b>noun</b>	<b>conj.</b>	<b>verb</b>

‘They removed the camp when it was morning.’

---

<sup>3</sup> There is only one exception, the marker of 3<sup>PL</sup> animate objects of transitive verbs. This marker is also one of the few morphemes that follow a nominative-accusative morphology, by contrast with the overall split intransitivity alignment of the language.

<sup>4</sup> Since °*tíde* is a bound root, it is not clear whether it should be glossed as an impersonal verb, like I did here, or as an intransitive verb ‘make a drumming sound’.

<sup>5</sup> The segmentation and glosses in all examples are mine (replacing the original word-by-word glosses from some sources), and the free translation is from the primary source whenever one is provided. Note that Dorsey’s free translation sometimes differs from the literary meaning of the sentence, in which case I add my own translation.

In sharp contrast with verbs, nouns have no dedicated morphology (except for inalienable possession markers on a restricted set of nouns, mainly kinship terms), but they regularly include verbal derivational morphology due to frequent conversions from verbs. Nouns can also be used predicatively, as in (1) where *hó<sup>n</sup>egó<sup>n</sup>ch<sup>h</sup>e* ‘morning’ is used as an impersonal predicate ‘to be the morning’.<sup>6</sup>

It seems that almost any lexeme of the language can be used as a predicate without a copula, and conversely verbs are frequently used as referential expressions in headless relative clauses. (This can later be seen in (9) or (55)). Umó<sup>n</sup>ho<sup>n</sup> has a series of eleven grammatical elements usually called “articles” because they often act as definite determiners after nouns. However, they are also used after verbs as auxiliaries, relativizers and evidential markers (Eschenberg 2005). Example (2) illustrates the use of articles as determiners after nouns and evidential markers after verbs.

(2) (Dorsey 1890: 410.19 / speaker: Ó<sup>n</sup>p<sup>h</sup>o<sup>n</sup>-to<sup>n</sup>ga)

<i>tó<sup>n</sup>de</i>	<i>k<sup>h</sup>e</i>	<i>no<sup>n</sup>-hó<sup>n</sup>ho<sup>n</sup> = biamá;</i>				
ground	the:HORIZ	INS:foot-tremble = PL.REPORT				
<b>noun</b>	<b>article</b>	<b>verb</b>				
<i>no<sup>n</sup>-tída = i</i>		<i>t<sup>h</sup>e,</i>	<i>hégazhi</i>	<i>amá:</i>	<i>Guuuu!</i>	
INS:foot-drumming.sound = PL		EVID	be.many	EVID	SOUND	
<b>verb</b>		<b>article</b>	<b>verb</b>	<b>article</b>	<b>onom</b>	

‘They made the ground tremble under their feet; they made a drumming noise as they ran in great numbers: *Guuuu!*’

For the reasons evoked above, some lexemes are variably translated by nouns or verbs in English (compare the translations of (14), (15) and (16)).

### 1.3 Umó<sup>n</sup>ho<sup>n</sup> data sources

This work is almost exclusively based on written resources, namely corpora and dictionary entries collected by Dorsey about 130 years ago (Dorsey 1890, 1891, 1892,

<sup>6</sup> Throughout this paper, I have standardized the spellings from the diverse sources, using the spelling of the Umó<sup>n</sup>ho<sup>n</sup> Language and Culture Center (ULCC), like other linguists (e.g. Rankin, 2008; Saunsoci & Eschenberg, 2016). The following letters and digraphs diverge from their IPA representation: <’> [ʔ]; <th> [ð]; <zh> [ʒ]; <sh> [ʃ]; <ch> [tʃ]; <j> [ç]; <x> [χ]; <o<sup>n</sup>> [õ] and [ã], which are allophones; <i<sup>n</sup>> [ĩ]. Expressive vowel length is shown by repetition of the vowel.

n.d.) and modern didactic materials (in their written form). In May 2023, I also conducted group elicitation sessions with elderly speakers and semi-speakers. They provided illustrative examples for the words that they were familiar with, examples of which are taken into account in this paper.<sup>7</sup>

As a result, apart from the few examples from my fieldwork, I have neither access to prosodic information nor to information on semantic subtleties beyond the (sometimes very short) definitions provided in the different sources. I have no negative evidence of the morphological and syntactic features of sound roots either.

Therefore, this paper is a philological study on a closed corpus. It investigates the Umó<sup>n</sup>ho<sup>n</sup> language as it is described in the existing documentation.

## 2. Data set

### 2.1 Dorsey's list of sound-imitating words

The starting point of this study is a publication by Dorsey in 1892 called “Siouan onomatopes”. This short paper (8 pages) mainly consists of an enumeration of Umó<sup>n</sup>ho<sup>n</sup> roots and lexemes that Dorsey considers to be onomatopoeia, which he defines as “word[s] or root[s] formed to resemble the sound made by the thing signified”. Dorsey presents, one by one, sound-imitating roots, generally followed by one or several derived or compounded forms, as in the following excerpt:

“*Tási* refers to a snapping sound, made by the aid of a rope, cord, or stiff hide; as, *thitási égo<sup>n</sup>*, to make such a sound by pulling a cord; *batási*, to make a snapping sound by punching against a rope or stiff hide.” (Dorsey 1892: 4)

Dorsey starts the definition of many roots with “used to describe the sound ...” or “denotes the sound of ...”, without clearly stating whether or not they are free forms. It seems that many of them are bound roots, since they only appear in texts and

---

<sup>7</sup> Many of the sound-imitating lexemes documented in Dorsey only came to us as word lists with rather short definitions, and no contextualized examples. This makes them particularly difficult to remember for speakers whose memory of the language is often triggered by the context of use. Additionally, I often didn't know how to pronounce the words accurately, since I did not have recordings of them, and sometimes no indication of stress placement. For these reasons, many sound-denoting lexemes documented in Dorsey could not be reused in elicitation.

dictionaries with some derivational process applied to them (affixation or reduplication). In this paper, I write these roots with an initial °.

Dorsey's (1892) data is very diverse overall. In the excerpt above, for instance, he presents the root °tási and two derived lexemes, *thitási* and *batási*. The first derived form is presented in a collocation, *thitási égo<sup>n</sup>* 'to make a snapping sound by pulling a cord' (see Section 3.5), while the other is provided alone. Dorsey also provides a few full sentences in his paper (e.g. entries 37 and 110 of Appendix A). Additionally, the roots provided belong to different parts of speech. Finally, Dorsey gives most of the definitions in a non-italicized font, but a few are in italics.

All the data provided in Dorsey (1892) is reproduced in Appendix A, alphabetically ordered by roots.<sup>8</sup> The roots are numbered in bold, from 1 to 53, and followed by any illustrative form that Dorsey added (derived lexeme, corresponding onomatopoeia or extra-linguistic sound imitated, collocation, sentence). Each form associated with a definition or a translation is an entry in Appendix A. Entries are numbered from 1 to 116.

In Dorsey (1892), the majority of the roots are displayed in two tables, ordered by the final syllable, and defined in the text, while a few are only mentioned in the body of the text. Dorsey thereby implicitly makes a distinction between a category of what I call "sound roots" and other sound-imitating lexemes. He also incorporates into his study a few roots which are not sound-imitating, but which he thinks could have been sound-imitating in an earlier stage of the language.

Before proceeding with an analysis of the formal and semantic features of the sound roots in Sections 3 to 5, I will present the roots that I have eliminated from the data set, and explain why, in Section 2.2.

## 2.2 Roots and lexemes eliminated from the data set

### 2.2.1 Bound roots that do not refer to sound

Dorsey refers to several roots that he suspects could have originally referred to a sound, but do not any more. He mentions a root °za attested in the compounded noun *no<sup>n</sup>bé ugáza* 'phalanges' (with *no<sup>n</sup>bé* 'hand'), and the root °mo<sup>n</sup>, which is found in a couple of verbs related to sharpening a scythe or an ax, and for which "the original reference (...) may have been to the sound made" (Dorsey 1892: 4). The latter root is also found in verbs completely unrelated to sound production, such as *bimó<sup>n</sup>* 'to knead

---

<sup>8</sup> Dorsey also provides a few cognates in other Siouan languages. These cognates are not taken into consideration in the present paper.

dough’. Additionally, he mentions the root °*dazhe* that creates verbs meaning ‘to chafe or blister’ and ‘to fillip with the fingers’, and °*duzhe* ‘split, cracked’ (Dorsey n.d.).<sup>9</sup> These verbs are probably mentioned because they are related to sound-denoting roots through the phenomenon called “consonant gradation” (Section 4.1). Since they are not sound-denoting themselves, I eliminated them from my data set.

Finally, Dorsey mentions the root *záde* which “conveys the idea of branching off or forking”, and provides example (3).

(3) (Dorsey 1892: 5 – my glosses)

<i>Hú</i>	<i>t<sup>h</sup>e</i>	<i>záde</i>	<i>ínahi<sup>n</sup></i> .
voice	the:VERT	fork	be.really.so

‘The voice is really indistinct – that is, the sound scatters instead of going straight to the person addressed.’

Example (3) refers to sound because of the subject *hú* ‘voice’, not because of *záde* ‘branching off’. As a consequence, I also eliminated the root *záde* from the data set.

### 2.2.2 Names of birds

At the end of his paper, Dorsey provides two bird names, which are referential and stand out from the rest of his data: *hák<sup>h</sup>ugthe* ‘whippoorwill’ and *káxe* ‘crow’. Cross-linguistically, many bird names are formed by an imitation of the songs or cries produced by the birds in question (Ullmann 1962: 86), and thus show iconicity: “a perceived resemblance between aspects of form and meaning” (Dingemanse 2019: 18). The two bird names do not display any morphological or syntactic particularity compared to other nouns, and they are not further studied.

### 2.2.3 Iconic verbs

The verbs *héch<sup>h</sup>i<sup>n</sup>* ‘to sneeze’ and *húxpe* ‘to cough’ are clearly sound-imitative. Except for their iconic nature, however, *héch<sup>h</sup>i<sup>n</sup>* and *húxpe* behave like regular verbs and have no morphological or syntactic specificity that would require a special treatment in a grammatical description. They can be inflected for person, as in (4) and (5). Like

---

<sup>9</sup> Although the action of splitting and cracking generally produces sound, the *meaning* of the root is not the sound, it is the action. This is a major difference with the sound roots analyzed in this paper.

°*dúzhe* ‘split, cracked’ mentioned above, they refer to an action that produces a sound, but their meaning is not restricted to the sound in question.

(4) (Saunsoci & Eschenberg 2016, 180 / speaker: Alice Saunsoci)<sup>10</sup>

**Tha-héch<sup>h</sup>i<sup>n</sup>** ki, í t<sup>h</sup>e ágaɣada = ga!  
**A2-sneeze** when mouth the:VERT cover = IMP.M  
 ‘Cover your mouth when you sneeze!’

(5) (Dorsey n.d., entry *húxpe*, my translation)

**hu < á > xpe**  
 < A1SG > cough  
 ‘I cough’

#### 2.2.4 Onomatopoeia

I define onomatopoeia as “the proper sound imitations [of sounds of extra-linguistic reality]” (Körtvélyessy 2020), like English *bang*, where the sound imitation both motivates and defines an onomatopoeia, and which are utterances of their own.<sup>11</sup> Two roots listed in Dorsey (1892) are onomatopoeia: the first one is *k<sup>h</sup>u* for the sound of a gun, illustrated in (6), and the second is the barking sound *hu-hu-hu* exemplified in Dorsey (1892) (see entry 29 of Appendix A). Note that none of them is written with a stress in Dorsey’s documentation.

(6) (Dorsey 1890: 436.6 / speaker: Páthi<sup>n</sup>-no<sup>n</sup>pázhi)

o<sup>n</sup>g-úthi’aga = í = de wa-k<sup>h</sup>ída = i: **K<sup>h</sup>u!** **k<sup>h</sup>u!** **k<sup>h</sup>u!**  
 A1PL-not.want = PL = as P1PL-shoot = PL **SOUND SOUND SOUND**  
 ‘When we refused to let them go, they shot at us: *Ku! ku! ku!*’

Syntactically, onomatopoeia share common features with interjections, since they do not enter into syntactic constructions with other parts of speech (Wilkins 1992: 124).

<sup>10</sup> Saunsoci & Eschenberg (2016) spell this verb *héchi*, but I follow Dorsey’s spelling here. They present it with the person prefixes on the left, as seen in (4), but Dorsey (n.d.) presents conjugated forms with person prefixes between the two syllables (A2: *he-thi-ch<sup>h</sup>i<sup>n</sup>*), like *hu-á-xpe* in (5). This suggests that, historically, these verbs are morphologically complex.

<sup>11</sup> With this definition restricted to “proper sound imitations”, verbs or nouns derived from onomatopoeia are not considered onomatopoeia any longer. For instance, sound imitation still motivates the derived verbal form *to bang*, but does not define it.



Several onomatopoeia are found in Dorsey's dictionary or texts, but interestingly, *k<sup>h</sup>u* 'bang' and *hu-hu-hu* 'woof' are the only ones so far which I have found used as predicates. In (7), *k<sup>h</sup>u* is the predicate of an impersonal construction. It takes no subject (either grammatically or conceptually), and as a consequence it is a holophrase. Note that in this instance (and only in this instance), it is stressed. See the discussion in Section 5.2 about the syntactic integration of onomatopoeia.

(7) (Dorsey 1890: 46.12 / speaker: Nudó<sup>n</sup>-axa)

*K<sup>h</sup>ú* = *biamá*.

SOUND = PP.REPORT

'The sound *ku* was made by shooting, they say.'

### 2.3 The sound roots

Out of the 53 roots presented in Dorsey's paper, and listed alphabetically in Appendix A, I have eliminated:

- 5 roots that do not refer to sound production
- 2 bird names which have an iconic phonetic form
- 2 verbs which have an iconic phonetic form
- 2 onomatopoeia

This leaves 42 sound-denoting verbal roots, called SOUND ROOTS in short. Like onomatopoeia, sound imitation both motivates their form and defines them, but unlike onomatopoeia they are not attested as utterances on their own. The next two sections discuss a number of features which are often associated with sound roots.

## 3. Morphological and syntactic features of sound roots

### 3.1 Generalities

All sound roots are verbal.<sup>12</sup> Some of them are independent (see Section 3.2), but many seem to be bound roots, as they are only attested in texts and dictionaries with some derivation. This seems, for instance, to be the case of the root °*táshi*, which is

---

<sup>12</sup> The definition of *tatáshi* in (8d) could correspond to a noun or a verb. This does not affect the classification of *tatáshi* as a verb, since verbs are easily used referentially, as mentioned in Section 1.2. See example (9).

attested with the most common derivations of sound roots, i.e. addition of an instrumental prefix and reduplication, as shown in (8). It is also derived with a suffix -é, which will be described in Section 3.3. Note furthermore that the root °*tashi* is not even defined by Dorsey, who only refers to the verbs derived from it.

(8) a. Root alone:

°*tashi* ‘seems to be used in two different ways: [*batáshi* and *gatashi*]’ (D92)

b. with instrumental prefixes (non-exhaustive list)

→ *gatáshi* ‘to make the sound *tshshsh* heard when one strikes a tree with an ax when the sap is flowing’ (D92)

→ *bitáshi* ‘to make a sound with the throat, almost “hawking”’ (DD)

c. with -é

→ *táshié-xti=o<sup>n</sup>* ‘said of persons, when they make much noise with their spoons, knives, or forks’ (DD)

d. with reduplication

→ *tatáshi* ‘said of the repeated ticking of a clock, or of the sounds made by many marbles’ hitting together’ [sic] (DD)

The prefixes *ga-* ‘by striking’ and *bi-* ‘by blowing’ illustrated in (8) are called “instrumental prefixes”, and they will be presented in Section 3.6.<sup>13</sup> All of the sound roots are attested with at least one instrumental prefix, and most are attested with several.

### 3.2 Syntactic functions of the independent roots

Independent sound roots are always intransitive verbs. They are sometimes used as the main predicate of the clause, as in (9) for the root *za’é* ‘to make an uproar (subject plural)’ (my translation), and in (10) for the verb *xtházhe* ‘to bellow, to scream out’. In (9), the verb is embedded in a relative clause. The plural proximate article *amá* acts as a relativizer, and the relative clause is in apposition to *gá-ama* ‘those ones’.

(9) (Dorsey 1890: 587.1 / speaker: George Miller)

“ <i>Ko<sup>n</sup>há,</i>	<i>gá-ama</i>	<i>za’é</i>	<i>amá</i>	<i>eát<sup>h</sup>o<sup>n</sup>=i</i>
grandmother.VOC	DEM-the:PX.PL	<b>make.uproar</b>	the:PX.PL	why = PL

<sup>13</sup> Instrumental prefixes should not be confused with the instrumental applicative marker which has no lexical meaning, and which adds an instrument as a verb object.

a,” á = *biamá*.

Q say = PP.REPORT

‘At last the Orphan said, “Grandmother, why do they make such a noise?”’

Literally: “Grandmother, those ones, those that make an uproar, why is it so?”  
said he, it is said.’

(10) (Dorsey 1890: 103.2 / speaker: Frank La Flesche)

*Mo<sup>n</sup>ó<sup>n</sup>u*                      *mo<sup>n</sup>thí<sup>n</sup>* = *biamá*,    *xthazhé*    *shti mo<sup>n</sup>thí<sup>n</sup>* = *biamá*.

paw.the.ground    walk = PP.REPORT    **bellow**    also walk = PP.REPORT

‘The blunt-horned Buffalo-bull kept (...) pawing the ground, and bellowing.’

They can also be used as verb modifiers in verb sequences, as exemplified in (54) for *za’é* ‘to make an uproar’. Verb sequences where the first verb modifies the second one are pervasive in the language. (See Marsault 2021: 140-142 for an introduction to the various types of verb sequences.)<sup>14</sup>

(11) (Dorsey 1890: 288.16 / speaker: Nudó<sup>n</sup>-axa)

*Za’é-xti*                                      *ó<sup>n</sup>he* = *hnó<sup>n</sup>* = *biamá*.

**make.uproar**-INTENS                      flee = HAB = PP.REPORT

‘They fled without exception, in great confusion.’ (i.e., making an uproar)

Note that by contrast with many bound sound roots, independent sound roots generally have an easily identifiable lexical meaning, and hence they are translated rather than glossed as SOUND.

By contrast with the iconic verbs presented in Section 2.2.3, sound roots are never attested with person prefixes (there is only one exception with *zúde* ‘to whistle’, discussed in Section 5.2). They are only attested with 3<sup>rd</sup> person subjects, which have no corresponding person prefix on the verb. It is not clear whether the absence of examples with 1<sup>st</sup> and 2<sup>nd</sup> person is due to a limitation of the corpus (the underived sound roots attested as verbs are rather scarce compared to the derived lexemes, and most of the texts are legends in the third person), or if it points to a morphosemantic restriction of the sound roots. In his dictionary, Dorsey presents *za’é* as a noun ‘a noise, hum, buzz, bustle,

---

<sup>14</sup> Although examples (10) and (11) contain serial verb constructions, they are different. Movement verbs such as *mo<sup>n</sup>thí<sup>n</sup>* ‘walk’ sometimes have a continuative meaning. Observe the difference of translations.

confusion' (despite its repeated attestations as a verb, as in (9)). The entry of the verb *xtházhe* 'to scream out' includes conjugated forms which are crossed out.

### 3.3 Derivation with the suffix -é

Several roots are attested with and without a final accented vowel -é of uncertain meaning. This suffix is only attested in Dorsey's dictionary (DD) and in one proper name listed in Fletcher & La Flesche (1911), and it is almost always associated with sound roots. As shown in (55), the bound root °*táxi* is mentioned in Dorsey (1892) as a bare root, but is only attested with -é in his dictionary.

- (12) a. °*táxi* 'formed from the sound *txxx*' (D92)  
 b. *táxié* 'the sound made in chopping wood in cold weather' (DD)  
 'knocking sound' (FLF:151), part of a proper name  
 c. *táxié t<sup>h</sup>ígthe* 'to make the sound *taxi* suddenly, in this place.  
 Applicable to a coyote or wolf when crunching bones.' (DD)

The definitions in (55) suggest that the root °*táxi* is the phonotactically correct way to represent an extralinguistic sound resembling *txxx*, but it has no lexical meaning; *táxié* is a lexeme describing the sound (maybe an impersonal verb); and *táxié t<sup>h</sup>ígthe* refers to an event (used in an intransitive construction).

Example (55) is the only example we have of a lexeme with -é in a sentence.

- (13) (Dorsey n.d.: entry *táxié* – my glosses and translation)  
*zho<sup>n</sup>*      *gáse = ma*      *táxi-é-xti = ó<sup>n</sup> = i*  
 wood      chop = the:OBV.PL      SOUND-e-INTENS = AUX = PL  
 Literally: 'those that cut wood produce the sound *táxi*'

In total, six sound roots are attested with the morpheme -é, including three which differ only by the point of articulation of their fricative (°*tasi*, °*tashi*, and °*taxi*). Examples in (14) through (16) reproduce some of the dictionary entries that include -é (in bold).

- (14) (Dorsey n.d.: entry *tásié t<sup>h</sup>ígthe*)  
*tási-é t<sup>h</sup>ígthe*  
 'to make the sound *tasi* suddenly, as in breaking a lariat, in this place'

- (15) (Dorsey n.d.: entry *p<sup>h</sup>ukie t<sup>h</sup>ígthe*)  
*p<sup>h</sup>uki-e t<sup>h</sup>ígthe*  
'a sudden sound, made by beating a soft robe, etc'
- (16) (Dorsey n.d.: entry *túshié*)  
*túshi-é-xti = o<sup>n</sup>*  
'said of the sounds of many distant reports of a gun,  
probable meaning: "they sound very distant"'

The meaning of *-é* is uncertain, apparently even for Dorsey himself (he writes "probable meaning" in (16)). It is not attested in modern documentation. Since it almost always follows sound roots, it could historically originate from the verb *é* 'to say', as suggested by Larson (2022). Note that cross-linguistically, many ideophones or sound-imitating lexemes are used with 'to say' (e.g., contributions in this volume by Authier, Bril, Meyer, Treis, Rose). The only example of *-é* on a non-sound-imitating root is given in (17).

- (17) a. *bazhú* 'callous: as any hard place on which the skin has formed by a burn or otherwise' (DD)  
'knotty, as wood that cannot be split' (DD)
- b. *bázhué t<sup>h</sup>ígthe* 'a knot or lump rises suddenly' (DD)

As seen above, the lexemes including a final *-é* are always followed by *t<sup>h</sup>ígthe* 'suddenly' or by *-xti = o<sup>n</sup>*, of uncertain meaning. Both *t<sup>h</sup>ígthe* and *-xti = o<sup>n</sup>* seem to imply an idea of suddenness, or of 'burst[ing] forcefully into experience' (an image suggested by Larson 2022).<sup>15</sup> It should be noted, however, that *t<sup>h</sup>ígthe* and *-xti = o<sup>n</sup>* are not triggered by the final *-é*, and can occur in other contexts. (See (20) and (34) for *t<sup>h</sup>ígthe*).

### 3.4 Construction with *t<sup>h</sup>ígthe* and *t<sup>h</sup>ítthe* 'suddenly'

A significant number of the roots listed in Dorsey (1892) are followed by *t<sup>h</sup>ígthe* or (in one instance) by *t<sup>h</sup>ítthe*, as illustrated in (18) and (19), respectively.

---

<sup>15</sup> Note that in other languages, ideophones sometimes convey suddenness in themselves (this is regularly the case in Thulung, described in Lahaussis this volume; see also Meyer and Quint this volume).

- (18) (Dorsey n.d.: entry *bthó<sup>n</sup>xe t<sup>h</sup>ígt<sup>h</sup>e*)  
*bthó<sup>n</sup>xe*    *t<sup>h</sup>ígt<sup>h</sup>e*  
 SOUND        suddenly  
 ‘to make a sudden crunching sound’<sup>16</sup>

- (19) (Dorsey n.d.: entry *xu’é át<sup>h</sup>iáthai*)  
*xu’é*        *át<sup>h</sup>iátha = i*  
 SOUND    pass.suddenly.PL = PL  
 ‘they (i.e. birds) passed with a sudden buzzing or roaring’

*T<sup>h</sup>ígt<sup>h</sup>e* and *t<sup>h</sup>ít<sup>h</sup>e* are compounds of *t<sup>h</sup>í* ‘to arrive here’, *gt<sup>h</sup>e* ‘to go back there’, and *thé* ‘to go there’, which encode third person plural and proximity with an initial *a-*, as observed in (19). Their semantics has evolved towards an expression of sudden action, whence my gloss ‘suddenly’. *T<sup>h</sup>ígt<sup>h</sup>e* is defined in Dorsey’s dictionary as ‘expressive of sudden action; used after other verbs’, but it is also attested in texts as ‘to become suddenly’ (following a noun or a verb expressing a property) and ‘to start suddenly’ (following a verb of action). *T<sup>h</sup>ít<sup>h</sup>e* is defined as ‘to begin, commence, start suddenly; to come forth, as an infant at birth’ in Dorsey’s dictionary, but is sometimes glossed ‘pass along’ or only ‘suddenly’ in texts. In (19), *t<sup>h</sup>ít<sup>h</sup>e* retains its original semantics of movement, with an addition of suddenness of action.

Dorsey’s dictionary includes at least 42 entries composed of collocations with *t<sup>h</sup>ígt<sup>h</sup>e* or *t<sup>h</sup>ít<sup>h</sup>e*, which suggests either that these collocations are frequent enough to deserve being added as a dictionary entry, or that they have a distinctive or expressive meaning. Interestingly, out of the 35 different roots represented in these entries (several entries with *t<sup>h</sup>ígt<sup>h</sup>e* involve the same root), 16 refer to the emission of a sound – almost half of them –, and 14 belong to the sound roots studied here. Fletcher & La Flesche (1911) also provide four proper names with *t<sup>h</sup>ígt<sup>h</sup>e* or *t<sup>h</sup>ít<sup>h</sup>e*, three of which involve sound emission. (The fourth involves movement.)

Three of the sound roots are only attested, in their underived form, in combination with *t<sup>h</sup>ígt<sup>h</sup>e*. The first one is *bthó<sup>n</sup>xe* ‘making a crunching sound’, a root generally poorly represented in the corpus. The second one is *tidé* ‘the sound heard in walking, striking a board, the ground’. The third one is the root *shkpáp<sup>h</sup>i* ‘splashing, slapping’,

---

<sup>16</sup> Example (18) is an unmarked verb (third person singular obviative), used as a citation form in the lexicon, and therefore translated as an infinitive.

only attested in a barely legible manuscript note, reproduced in (20) (XX and vowels in brackets indicate illegible text). As we see, it is followed by *égo<sup>n</sup>* and *t<sup>h</sup>igthe*.

(20) (Dorsey n.d.: entry *thashkáp<sup>h</sup>i*)

<i>shkáp<sup>h</sup>i</i>	<i>égo<sup>n</sup></i>	<i>át<sup>h</sup>[ia]gthai?</i>
SOUND	like	appear.suddenly
‘XXX be gone with a splash XX’		

This data suggests that *t<sup>h</sup>igthe* and *t<sup>h</sup>ithe* could act as light verbs in these contexts, which would mean that the preceding roots (*bthó<sup>n</sup>xe*, *tidé*, *shkáp<sup>h</sup>i*) have a limited predicative function. These roots are considered free roots, but more data is needed to confirm that they can be used as verbs on their own.

### 3.5 Construction with *égo<sup>n</sup>*

The lexeme *égo<sup>n</sup>* is multifunctional and very frequent. Historically, it is composed of the demonstrative marker *é* and the morpheme *-go<sup>n</sup>*, which has a similitive meaning. It is found in many different syntactic constructions, and with slightly different meanings. We can identify at least the following uses:<sup>17</sup>

1. Verb ‘to be so’, ‘to be like’ (conjugates with 1<sup>st</sup> and 2<sup>nd</sup> patientive)
2. Verb ‘it is (somewhat) like it’ (always 3<sup>rd</sup> person, does not conjugate)
3. Conjunction ‘as, having’, after a dependent clause

The three functions are exemplified in examples (21), (22) and (23), respectively. We see in (22) and (23) that *égo<sup>n</sup>* does not conjugate. This shows that it behaves differently in different syntactic constructions, hence the need to distinguish the three functions.

(21) (Dorsey 1890: 731.10 / speaker: Mo<sup>n</sup>ch<sup>h</sup>ú-No<sup>n</sup>ba)

<i>Zé &lt; thi &gt; tha = í</i>	<i>ki, díxe</i>	<i>é &lt; thi &gt; go<sup>n</sup> = bázhi</i>	<i>ta = í.</i>
< P2 > doctor = PL	if be.scabby	< P2 > be.so = NEG.PL	IRR = PL

<sup>17</sup> The different meanings and functions of *égo<sup>n</sup>* are obviously linked. There are attested diachronic pathways from simile (here, ‘to be like’) to subordinators (Heine & Kuteva 2002: 273-274; Güldemann 2008: 317ff).

‘If you are vaccinated you will not have the small-pox.’

Literally: ‘If they doctor you, you will not be scabby (= with smallpox).’

(22) (Dorsey 1890: 152.18 / speaker: Nudó<sup>n</sup>-axa)

*To<sup>n</sup>dé,*                      *u < thí > nadátho<sup>n</sup>*      **égo<sup>n</sup>,**                      *á = biamá.*  
 daughther’s.husband      <P2> used.to.heat      **it.is.like**                      say = PP.REPORT

‘O daughter’s husband, have you become somewhat accustomed to the heat?’

(23) (Dorsey 1890: 87.14 / speaker: Nudó<sup>n</sup>-axa)

*Hau!*      *u < thí > tha = i*                      **égo<sup>n</sup>**      *wi-nó<sup>n</sup>o<sup>n</sup>*                      *pí*                      *ha,*  
 ho!      <P2> talk.of = PL      **as**      A1SG:P2-hear      A1SG.come.here      DECL.M  
*á-biamá.*

say = pp.report

‘Well, as you have been reported (= famous), I have been coming to hear you.’

The sound roots are regularly associated with *égo<sup>n</sup>*, apparently with the second meaning ‘it is (somewhat) like it’. Example (24) shows how the speaker Arlington Saunsoci spontaneously added *égo<sup>n</sup>* after the reduplicated sound root *zuzúde* ‘whistling sound’ during an elicitation session on sound imitations.

(24) (Fieldwork session May 11, 2023: 5’39-6’03 / speaker: Arlington Saunsoci)

AS: *Zuzúde wa’ó<sup>n</sup>*. ‘She’s whistling the song.’

JM: And have you ever heard *zuzúde* used alone? without *wa’ó<sup>n</sup>*? Like just *zuzúde*, or *zuzúda*? Does it sound... natural to you?

AS: If you say, *zuzúd(e)-ego<sup>n</sup>*, it’s kind... you’re saying: ‘it’s kind of a whistle’

Almost all of the examples with *égo<sup>n</sup>* ‘it is (somewhat) like it’ involve sound roots derived with instrumental prefixes (more on this in Section 3.6). In a dictionary entry reproduced in (25), Dorsey provides a conjugated form for the phrase *mútaxi égo<sup>n</sup>*, ‘to make the sound *taxi* (...)’. As can be seen, only the verb *mútaxi* is conjugated, but not *égo<sup>n</sup>*.

(25) (Dorsey n.d.: entry *mútaxi*, my glosses)

*mútaxi égo<sup>n</sup>*

to make the sound *taxi* by firing a gun and letting the hammer fall; to make this sound by shooting and hitting a bone



[A2 form:] *mú-tha-taxi*                      *égo<sup>n</sup>*  
 INS:shoot-A2-SOUND    it.is.like

Additionally, Dorsey does not provide any definition for *mútaxi*. Instead, he directly jumps to the sub-entry *mútaxi égo<sup>n</sup>* in (25), as if *mútaxi*, although identifiable as a lexeme, could not be used without *égo<sup>n</sup>* ‘it is (somewhat) like it’. This is similar to example (24), and also reminiscent of the dictionary entries that include *t<sup>h</sup>ígthe* and *t<sup>h</sup>ítthe* ‘suddenly’, in Section 3.4.

In total, thirteen of the 42 sound roots are attested with a *égo<sup>n</sup>* ‘it is (somewhat) like it’, most of the time in combination with an instrumental prefix.

### 3.6 The instrumental prefixes on the sound roots

Umó<sup>n</sup>ho<sup>n</sup> (like other Siouan languages) has a series of prefixes which specify how the action is carried out, and that are usually called “instrumental” in the literature on Native North American languages. Table 1 presents the nine instrumental prefixes of Umó<sup>n</sup>ho<sup>n</sup>, and the main meanings for each of them. (See UNL-OLIT 2018:433-63 and Marsault 2021:289-304 for a more detailed description and semantic analysis.)

Instrumental prefixes mostly combine with verbs and bound verbal roots, and they invariably create verbs, which I call “instrumental verbs”. As mentioned in Section 3.1, all sound roots are attested in combination with at least one instrumental prefix.

Prefix	Meanings
<i>ba-</i>	by pushing
<i>bi-</i>	by pressing; by blowing with the mouth
<i>ga-</i>	by striking; by sudden action of wind; by falling
<i>má-</i>	by cutting; with a blade
<i>mú-</i>	by shooting; with a stream of water or wind; with shooting sensations (like pain)
<i>ná-</i>	by extreme temperature
<i>no<sup>n</sup>-</i>	with the feet; under its own power
<i>tha-</i>	with the mouth
<i>thi-</i>	by pulling; with the hands

**Table 1:** Main meanings of the instrumental prefixes

Instrumental prefixes generally have a causative function. For instance, the prefixes *thi-* ‘with the hands’ and *ga-* ‘by striking’ have a causative function in (26) and (27), respectively. In each case, they create transitive verbs which are the head of independent clauses.

(26) (Saunsoci & Eschenberg 2016: 140 / speaker: Alice Saunsoci)

*Tizhébe*      *t<sup>h</sup>e*                      ***thi-’áxa***.  
 door            the.VERT            **INS:hand-squeak**  
 ‘He squeaked the door (made the door creak).’

(27) (Dorsey 1890: 266.10 / speaker: Te-úko<sup>n</sup>ha)

*tí-ha*                      ***ga-p<sup>h</sup>úki = biamá***.  
 tent-skin              **INS:force-SOUND = PP.REPORT**  
 ‘They made the tent skins sound by hitting them.’

Quite often, however, the instrumental verbs derived from sound roots are found in different syntactic constructions. In (28), the main verb of the clause is *uxpátha* ‘to fall’. The sound root *kúge* ‘hollow sound’ (ULCC), derived with *ga-* ‘by falling’, creates the verb *gakúge* which acts as a modifier of *uxpátha* ‘to fall’. Here, the meaning of the prefix *ga-* ‘by falling’ is redundant with the main verb of the clause.

(28) (UNL- OLIT 2018: 562)

*Kúge*            *tho<sup>n</sup>*            ***ga-kúge***            *uxpátha*.  
 drum            the:RND      **INS:fall-SOUND**    fall  
 ‘The drum fell with a thump.’

In example (29), the use of the instrumental prefix *tha-* ‘with the mouth’ on the sound root *sathú* ‘to rattle’ is redundant with the same prefix on another verb, and corresponds to an echo construction (an example of “echo phenomena”, often used in the description of associated motion markers; see Guillaume 2009, Jacques 2023). This sentence is provided by Dorsey in lieu of a definition for the verb *thasáthu*. His translation suggests that *thasáthu* is used intransitively.<sup>18</sup>

<sup>18</sup> Primary stress must fall on one of the first two syllables of the word. This is why the stress shifts to the left when the prefix *tha-* ‘with the mouth’ is added to the root *sathú* ‘to rattle’.

(29) (Dorsey n.d.: entry *thasáthu*)

*Wat<sup>h</sup>ó<sup>n</sup>zi*    *tha-shpí*                    *that<sup>h</sup>á=i*    *ki*,            *tha-sáthu=i*                    *ha*  
 corn            INS:mouth-off            eat = PL    when    INS:mouth-SOUND = PL DECL.M  
 ‘to make the sound heard when corn is eaten from the cob’  
 Literally: ‘When they eat corn by biting it off [from the cob], they produce the sound *sathu* with their mouths’

The same can be said of the verb *thixú’e* ‘to make the sound *xu’é* by tearing’ in (30). Again a sentence is provided instead of a definition in a dictionary entry.

(30) (Dorsey n.d.: entry *thixú’e*)

*Thi-btháza=i*                    *ki*,    *hú*            *t<sup>h</sup>e*    *thi-xu’é*.  
 INS:hand-open = PP    if    sound    the    INS:hand-SOUND  
 ‘If any object is torn, the sound is *thixu’é*.’

We see in (31) another example of *thixú’e* ‘to make the sound *xu’é* by tearing’ in a narrative. Again, it is used intransitively, following another verb with *thi-* ‘with the hands’, in a clause that modifies this other verb.

(31) (Dorsey 1890: 259.11 / speaker: Te-úko<sup>n</sup>ha)

*égithe*            *te-néxe*                    *thi<sup>n</sup>k<sup>h</sup>é*                    *thi-btháza=biamá*,  
 at.length    buffalo-bladder    the:OBV.SIT    INS:hand-open = PP.REPORT  
*thi-xú’(e)*                    *égo<sup>n</sup>=ma*.  
 INS:hand-SOUND    it.is.like = EVID  
 ‘At length the Buffalo-bladder was torn open, making the sound *xu’é*.’<sup>19</sup>

We observe that in (31), the instrumental verb *thixú’e* is followed by *égo<sup>n</sup>* ‘it is (somewhat) like it’. Indeed, many instrumental verbs built on sound roots are followed by *égo<sup>n</sup>*. Example (25) in Section 3.5 showed that Dorsey sometimes provides dictionary entries for an instrumental verb alone, and then only defines or exemplifies it with a sequence with *égo<sup>n</sup>*. Example (32), an extract of my fieldwork with current-day speakers, also suggests that some instrumental verbs are mainly or only used with *égo<sup>n</sup>*.

<sup>19</sup> Dorsey writes *thixú’egó<sup>n</sup>-ma* as a single word. For the sake of uniformity, I write *égo<sup>n</sup>* as a separate word in this example.

(32) (Fieldwork session, May 10, 2023 / speakers: Arlington Saunsoci and Dwight Robinson)

JM: Any other word with *tushi*? Like *tutúshi*, or *thitúshi*... No? Only *gatúshi*?

[various people mutter to themselves, trying to remember]

DR: *gatúshi* is the only one that I know.

AS: *thitúsh' égo<sup>n</sup>*. You say that when you say ‘making a popping noise’.

In (33), we see the entry *gap<sup>h</sup>úki égo<sup>n</sup>*, created in addition to the entry *gap<sup>h</sup>úki*. Dorsey provides a sentence which does not include the entry *gap<sup>h</sup>úki égo<sup>n</sup>*, but rather describes how its sound is produced (by hitting robes). He then specifies that *gap<sup>h</sup>úki égo<sup>n</sup>* “refers to the sound”, while the verbs *gako<sup>n</sup>* or *gako<sup>n</sup>ko<sup>n</sup>* “refer to the purpose or effect” (the latter are not cited in Dorsey 1892 and are not built on a sound root). This suggests that *gap<sup>h</sup>úki égo<sup>n</sup>* is a more depictive way to refer to the event.

(33) (Dorsey, n.d.: entry *gap<sup>h</sup>úki égo<sup>n</sup>*, my glosses)

*gap<sup>h</sup>úki égo<sup>n</sup>*, explained thus by [Wajepa] (1889):

*waí<sup>n</sup> t’o<sup>n</sup> shó<sup>n</sup> ut<sup>h</sup>t<sup>n</sup>.*

robe there.is(?) so(?) hit

This phrase refers to the sound, but *gako<sup>n</sup>* or *gako<sup>n</sup>ko<sup>n</sup>* refer to the purpose or effect.

Finally, example (34) illustrates an instrumental verb followed by *t<sup>h</sup>íthe*. Unfortunately, there is no example of this verb in a sentence, but it seems again that the instrumental verb still primarily refers to the sound produced. The entry *gakátho<sup>n</sup>* does not exist.

(34) (Dorsey n.d.: entry *gakátho<sup>n</sup> t<sup>h</sup>íthe*)

*ga-kátho<sup>n</sup> t<sup>h</sup>íthe*

INS:shake-SOUND suddenly(?)

‘to sound or rattle, as an old kettle containing stones, etc., when shaken’

It seems that the semantic and syntactic features of the sound roots are often preserved when these roots are derived with instrumental prefixes. The resulting instrumental verbs refer to sound emission above all (and not to some action resulting in a sound emission), and Dorsey seems to have difficulties to define them (preferring

direct examples to definitions). They are often used as verb modifiers or in adverbial clauses, and we find them in the collocations with *t<sup>h</sup>ígthe/t<sup>h</sup>ítthe* ‘suddenly’ or *égo<sup>n</sup>* ‘it is (somewhat) like it’.

In total, I have found evidence of unusual collocations (with *t<sup>h</sup>ígthe* or *égo<sup>n</sup>*), valency (intransitive), or semantic value of the prefix (redundancy) in verbs derived from 17 different roots.

#### 4. Form-meaning mappings

This section describes the form-meaning mappings of sound roots. A large number are subject to the phenomenon called “consonant gradation” that will be described in Section 4.1. In Section 4.2, I argue that sound roots are distinct from onomatopoeia, although both are sound-imitative.

##### 4.1 Consonant Gradation

Siouan languages are known for exhibiting a peculiar type of iconicity affecting fricatives, sometimes called “consonant gradation”, where different points of articulation (alveolar, post-alveolar, velar) symbolically refer to different grades, e.g. of intensity.<sup>20</sup> This phenomenon is salient enough to be mentioned in Mithun’s (1999) survey of Native languages of North America, in a section dedicated to sound symbolism. Its most emblematic example is the three-grade color distinction exemplified in (4) for Umó<sup>n</sup>ho<sup>n</sup>.<sup>21</sup>

(35) (Dorsey n.d.)

<i>zí</i>	↔	<i>zhí</i>	↔	<i>ɣí</i>
‘yellow’		‘orange-red’(?)		‘brown’

This feature probably already existed in Proto-Siouan (Matthews 1970; Rankin et al. 2015), but is no longer productive or is only “semi-productive” (Matthews 1970) in

---

<sup>20</sup> Consonant gradation in Siouan is completely different from systematic consonant alternations triggered by phonological or morphological contexts as found in other languages such as Finnish. These systems are sometimes also called “consonant gradation” (see Merrill 2018: 29).

<sup>21</sup> The Lakota cognates *zí* ‘yellow’, *ží* ‘tawny’, and *ǵí* ‘brown’ (API: *zí*, *ží*, and *ɣí*) are one of the sets most often cited in the literature.

the daughter languages. In Umó<sup>n</sup>ho<sup>n</sup>, most examples involve pairs of roots rather than triplets. The points of articulation in contrast vary.

Many sound roots are subject to consonant gradation. The semantic relation between formally related sound roots is often difficult to assess. In a number of cases, the further back the point of articulation is, the more intense or hoarse the imitated sound, as in (36).

- (36) a. *zúde* ‘a whistling sound’ (D92) ↔ *zhúde* ‘the expulsion of the breath by a person or animal that is nearly exhausted from running, etc.’ (D92)
- b. *no<sup>n</sup>táshi* ‘the sound of birds walking on a hard surface with their nails hitting against it: *Ts! Ts! Ts! Ts! Ts!*’ (SLW) ↔ *no<sup>n</sup>táxi* ‘refers to the sound of a horse’s feet on hard, but not frozen, ground’ (D92)

In other instances, the semantic distinction seems to have been neutralized. For instance, Dorsey (n.d.) does not provide any definition for *táshié t<sup>h</sup>ígthe*, but mentions that it is synonymous with *táxié t<sup>h</sup>ígthe* ‘to make the sound *taxi* suddenly, in this place’. Also, note that in (36b), the definition of *no<sup>n</sup>táshi* (with post-alveolar fricative) includes *ts* (alveolar) as an imitation of the sound produced by birds.

In still other cases, the semantic relationship between two roots is difficult to assess because each of them has several meanings. The roots *su’é* and *xu’é*, for instance, are described in Dorsey (1892) as referring to a variety of sounds (see Appendix A), and a few more sounds are listed in ULCC (2018). The latter source explicitly states that “*xu’é* is a sound word” and that “*xu’é* is softer than *su’é*” (ULCC, 2018: 36), which is contrary to the examples in (36).

Finally, several sound roots enter into a consonant gradation relationship with other roots which do not always refer to sound. For instance, *bthó<sup>n</sup>xé* ‘crunching sound’ is linked to *bthó<sup>n</sup>ze* ‘fine, as hair, silk, flour, etc’, attested in Dorsey’s dictionary (Dorsey n.d.).<sup>22</sup>

Among the 33 sound roots that have a fricative in my data set, 20 are involved in consonant gradation with another root, that is, more than 60%. They combine an imitative type of iconicity (sound imitation) with a relative type of iconicity (consonant gradation) (see Johansson et al. 2020 for a typology of iconicity). Consonant gradation is by no means restricted to sound roots, however. In Marsault

<sup>22</sup> Another example will be mentioned in Section 5.3.

(accepted), I identify 106 roots which enter into consonant gradation relationships, among which we find 23 sound-denoting roots. Thus, sound is involved in only 21% of all roots with consonant gradation, a percentage that makes it nonetheless the main semantic domain represented.

## 4.2 Sound roots vs. onomatopoeia

In Section 2.2.4 I defined onomatopoeia, and I eliminated two onomatopoeia cited in Dorsey (1892) from my data set. In addition to the fact that, to the best of my knowledge, none of the features described in Sections 3 and 4.1 apply to any onomatopoeia, a few examples show that the same sound can be referred to by a sound root and an onomatopoeia completely different from one another. This is the case in (37), with a rattling sound expressed with the sound root *sathú* and the onomatopoeia *ch<sup>h</sup>uuuu*.

- (37) (Dorsey 1890: 411.1 / speaker: Ó<sup>n</sup>p<sup>h</sup>o<sup>n</sup>-to<sup>n</sup>ga)  
*thi-p'ó<sup>n</sup>de*      *gó<sup>n</sup>*      *thi-sáthu = hnó<sup>n</sup> = biamá:*      *ch<sup>h</sup>uuuu*.  
 INS:NEU-shake    as      INS:NEU-SOUND = HAB = PL.REPORT      SOUND  
 ‘Whenever he lifted his tail, he rattled it: *Chuuuu* (whispered).’

Another example can be seen in (2), with the sound root °*tíde* and the onomatopoeia *guuuu* for a drumming sound. Finally, note that the fire of a gun can be referred to by the onomatopoeia *k<sup>h</sup>uuu*, as exemplified in (6), with the sound root *túshi*, exemplified in (16), and with the reduplicated *ch<sup>h</sup>ich<sup>h</sup>ízhe*. All of them are listed in Dorsey (1892) and can be found in Appendix A.

## 5. Sound roots as a lexical class

### 5.1 Are sound roots ideophones?

Dingemanse (2019: 16) proposes the following cross-linguistic definition of the ideophone, to be used for comparative and typological purposes: “a member of an open lexical class of marked words that depict sensory imagery”. One of the main difficulties in analyzing the Umó<sup>n</sup>ho<sup>n</sup> data is to assess how “marked” sound roots are, compared to the rest of the lexicon. Are they salient enough to justify being identified

as a distinct lexical class within the class of verbal roots? In Section 5.1.1, I review the features described in Sections 3 and 4 and discuss the markedness of sound roots. Then in Section 5.1.2, I turn back to Dingemanse's definition and discuss how well the sound roots correspond to it.

### 5.1.1 The markedness of sound roots

In Section 3.3 through Section 3.6, I identified four morphological and syntactic features which mark some sound roots as distinct from the rest of the lexicon, and in Section 4.1, I presented consonant gradation as another possible marked feature. These features are numbered from 1 to 5 in the list below. We see that the markedness of consonant gradation is weak, due to the high number of roots involved in consonant gradation which are not sound roots.

1. morphological construction involving the suffix *-é* (see Section 3.3)
  - roots involved: 6 sound roots / 1 other root
2. collocation with *t<sup>h</sup>igthe* '(to become) suddenly' or *t<sup>h</sup>ithe* '(to become/pass) suddenly' (see Section 3.4)
  - roots involved: 13 sound roots / 2 sound-denoting roots not mentioned in Dorsey 1892 / 20 other roots
3. collocation with *égo<sup>n</sup>* 'it is (somewhat) like it' (see Section 3.5)
  - roots involved: 13 sound roots / no available figure for other roots
4. morphosyntactic construction where the instrumental prefixes lose their causative function (see Section 3.6)
  - roots involved: 17 sound roots / no available figure for other roots
5. consonant gradation
  - roots involved: 20 sound roots / 3 sound-denoting roots not mentioned in Dorsey (1892) / 83 other roots

The features attested for each sound root are listed in the table in Appendix B. This table shows that the roots differ widely in the number of features they are attested with. Nine roots are attested with 3 or 4 of the morphological and syntactic features, barring consonant gradation. By contrast, twenty roots are found with none of these features.



In Sections 2.3 and 3.1, I defined sound roots as verbal roots which refer to sounds, can be derived with at least one instrumental prefix, and are not attested with person prefixes. This is a rather broad definition, in which the semantics of the root is the only criterion that identifies the sound roots among the numerous verbal roots that combine with instrumental prefixes.

The morphological and syntactic features described in Sections 3.3-3.6 only concern about half of the sound roots, and to various degrees. Thus, there is a certain degree of morphological and syntactic markedness, but the number of roots to be considered as marked varies depending on which features or combination of features are considered relevant. It should also be noted that there is a great disparity in the number of tokens attested for each root. Among the twenty roots which have none of the morphological or syntactic features above, twelve are never attested in narratives nor in example sentences.<sup>23</sup> This makes it almost impossible to know if, for instance, the instrumental prefixes have a causative function or not (dictionary definitions not being specific enough). It also reduces the possibility to find them used in special collocations. A few of them, like °túp<sup>h</sup>i ‘pattering sound’, are not even attested in Dorsey’s dictionary, so its mention in Dorsey 1892 is the only attestation in all the available documentation. It is possible that narratives and letters, the two text types recorded by Dorsey, are not ideal to illustrate sound roots, compared to interactions. The modern didactic materials possibly miss sound roots, too. They are often built on translations from English to Umó<sup>n</sup>ho<sup>n</sup>, and the non-prevalence of sound-denoting lexemes in English may be an obstacle to their documentation and transmission in a context of language shift.

### 5.1.2 Comparison to Dingemane’s definition of ideophones

Dingemane’s definition of ideophones is made up of five criteria: (a) open class membership; (b) conventionalization as words; (c) phonological, morphological or syntactic markedness;<sup>24</sup> (d) depictive value; (e) meaning in the domain of sensory imagery. I will discuss them one by one below.

---

<sup>23</sup> The twelve roots that appear in no narrative nor text example, either alone or in a derived form, are: °ch<sup>h</sup>áki, °dáze, °k’éxe, °s’ú, °shathú, °shka, °shtáki, °skáp<sup>h</sup>i, °túp<sup>h</sup>i, °xátho<sup>n</sup>, xthó<sup>n</sup>zhe, and °zide.

<sup>24</sup> Dingemane (2019:15) writes “ideophones are MARKED, i.e. they have structural properties that make them stand out from other words”, without specifying what he means by “structural properties”. I interpret it to broadly means any kind of phonological, morphological, or syntactic marking.

## (a) open class membership

This is a feature difficult to study in the current state of documentation of the language. I have identified 42 sound roots in the data presented in Dorsey's paper, although this number decreases if we only take into account those that have marked morphological or syntactic features. This makes it a rather small class, but attested cases of polysemy, semantic shifts and interpersonal variation attested suggests that it is an open class (see Sections 4.1, 5.3, and comment on the root *k'úshi* in Appendix B).

## (b) conventionalization as words

Sound roots are conventional items. They respect Umó<sup>n</sup>ho<sup>n</sup> phonology and phonotactics (by contrast with imitations of extra-linguistic sounds, like *txxx* in (55)), and they have meaning, as we have seen in dictionary definitions. The difficulty to gloss many of them is due to the absence of equivalent terms in English. However, many of them are not words, in the sense that they are bound roots.

## (c) phonological, morphological or syntactic markedness

The markedness of the sound roots is discussed in Section 5.1.1. A core of nine to 22 roots are morphologically and/or syntactically marked, while the others are not.

## (d) depictive value

The sound roots are not completely depictive, since pure depiction can only apply to units which are not grammatically integrated (see Section 5.2). The punctuation and transcription used in the written sources do not suggest any prosodic foregrounding or expressive lengthening of vowels in the original oral form, by contrast with onomatopoeia, for instance. But like in onomatopoeia, sound imitation defines sound roots, which makes them partly depictive.

## (e) sensory imagery

As expressions of sounds, the sound roots come under the semantic domain of sensory imagery.

As a summary, the sound roots form a set of conventional lexical items, possibly open to new members, whose meaning relates to the sensory imagery, and whose sound-denoting nature makes them more depictive than other verbal roots. They do not fit Dingemans's comparative concept of ideophones, however. They are not

systematically marked, and they are never as depictive and prosodically foregrounded as onomatopoeia are.

Sound roots can be considered one of the ideophone-like phenomena whose study informs us about ideophone systems cross-linguistically, by exploring their boundaries. Sound roots show dynamics similar to ideophones. Indeed, the continuum of syntactic integration and the semantic extension towards other sensory domains is attested for sound roots as it is for ideophones, as I will show in Sections 5.2 and 5.3, respectively.

## 5.2 A continuum of grammatical integration

Studies show that ideophones tend to gradually lose their depictive force as they become more integrated into the grammar (Dingemanse & Akita, 2016; Dingemanse, 2017). This also applies to Umó<sup>h</sup>o<sup>n</sup> onomatopoeia and sound roots, and can be represented as a continuum.

Table 2 illustrates how eight different roots, in different collocations and with different derivations, extend from completely depictive on the left, to completely descriptive on the right. At one extreme of this continuum are onomatopoeia, which are depictive and never syntactically integrated. (They are featured in dark gray cells.) At the other extreme are nouns or verbs whose meaning is not sound-denoting, although they are derived from sound roots. (They are featured in light gray cells.) The examples in between include sound roots in diverse morphological and syntactic constructions, as long as they preserve their sound-denoting meaning. The medium gray cells also include verbs ‘to bark’ derived from the onomatopoeia *hu-hu-hu* ‘woof’.

We observe that the examples broadly extend on a diagonal from the top left (depictive) to the bottom right (descriptive). The first one, *xwiii* ‘sounds of tree falling’, is a typical example of an onomatopoeia. It appears only once in Dorsey’s published texts, as a purely depictive item. The onomatopoeia *k<sup>h</sup>u* ‘bang’ on the second line is also used predicatively, as shown in (7). As an impersonal predicate, it remains relatively free, since it is not linked to a subject or an object.

The onomatopoeia *hu-hu-hu* and lexemes derived from it extend on the third line. To the best of my knowledge, it is the only onomatopoeia that is converted into verbs and nouns. It is a verb modifier in (38), and a verb on its own in (39).

- (38) (Dorsey, n.d.: entry *hu í*)<sup>25</sup>  
*Hú shi a? K<sup>h</sup>e! í=ga*  
**bark** A2.come.here Q well come.here = IMP.M  
 ‘Are you coming barking? Well! Be coming! (I will hit you)’

By contrast with the onomatopoeia *k<sup>h</sup>u* ‘bang’ used predicatively in (7), the verb *hú* ‘to bark’ below is intransitive. It is semantically and syntactically linked to the subject ‘the wolf’.

- (39) (ULCC, 2018: 10 / speakers: Marcella Woodhull Cayou & Donna Morris Parker)  
*Shó<sup>n</sup>to<sup>n</sup>ga ak<sup>h</sup>a hó<sup>n</sup>=no<sup>n</sup>-di hú gthí<sup>n</sup>=no<sup>n</sup>=biama.*  
 wolf the:PX.SG night = HAB-LOC **howl** sit = HAB = PP.REPORT  
 ‘The wolf howls at night, they say.’

Finally, *hú* is also a noun ‘voice’, ‘sound’, in which case it becomes purely descriptive and referential.<sup>26</sup>

The remaining lines of Table 2 illustrate how five sound roots are more or less grammatically integrated. As previously mentioned, these roots are never attested as onomatopoeia, and they are never as depictive as onomatopoeia are.<sup>27</sup>

The examples on each line are roughly organized from the least to the most grammatically integrated. I consider that the collocations with *égo<sup>n</sup>* and with *t<sup>h</sup>ígthe* feature the same degree of grammatical integration, which is relatively low. Verbs are more integrated, and I consider that their grammatical integration increases at the same time as their valency, because they are linked to more arguments.

In Section 5.1.1, I analyzed the sound roots markedness by looking at the morphological and syntactic constructions each root is attested in. Although this is a necessary step in the identification of sounds roots as a lexical class, it sometimes brings contradictory results. Dingemanse (2017) documents how the same Siwu

<sup>25</sup> Note that the sequence *hu í* ‘to come barking’ has its own entry in Dorsey’s dictionary. The movement verb *í* ‘to come here’ has an irregular conjugation.

<sup>26</sup> Headman & O’Neil (2019) indicate a difference of vowel length between the noun and the verb in Pá<sup>n</sup>ka: *hú* ‘voice’ vs. *húu* ‘to howl’.

<sup>27</sup> The sound root *°táxi* imitates the sound *txxx* or the sound *t’x t’x hyui*, according to Dorsey (1892), but it is not clear whether these forms are actually onomatopoeia or not (considering onomatopoeia as conventional items, and not improvised imitations of extra-linguistic sounds). Even if they were, *txxx*, *t’x t’x hyui*, and *°táxi* cannot be considered the same root.

ideophones can be used as “pure”, depictive ideophones, or as grammaticalized (and de-ideophonized) ones, following Dwyer & Moshi’s (2003) distinction. The same variation of expressiveness can be observed in Umó<sup>n</sup>ho<sup>n</sup>. Beyond the various degrees of grammatical integration where the sound roots are still sound-denoting, illustrated in the medium gray cells of Table 2, we observe categorial changes when the root loses its sound-denoting meaning.

Dorsey (1892) presents the root *zúde* as ‘denot[ing] a whistling sound, such as a man makes’. Then, he presents the phrase *zuzúde wa’o<sup>n</sup>* for ‘to whistle’ (with the reduplicated root *zuzúde* ‘whistling sound’ used as the modifier of the verb *wa’o<sup>n</sup>* ‘to sing’), and the verb *gazúzude* ‘to roar or whistle often, as the wind does’.

Dorsey’s dictionary also contains one entry for *zud(e)-égo<sup>n</sup>*, ‘with a whirr (of the wings)’. From these examples, we can classify *zúde* ‘whistling sound’ as a sound root, because the sound imitation constitutes its meaning, it can be derived with an instrumental prefix, it is used as a verb modifier and it is attested with *égo<sup>n</sup>* ‘it is (somewhat) like it’.

However, *zúde* is also a verb ‘to whistle’ in several dictionaries (DD, ST, SLW). In this case, it can take person prefixes and primarily refer to the action of whistling. It can be used in non-declarative-affirmative sentences, as in (40).

- (40) (UNL-OLIT 2018: 527)
- |                    |                          |           |
|--------------------|--------------------------|-----------|
| <i>Tha-zúde</i>    | <i>u &lt;thá&gt;kihi</i> | <i>a?</i> |
| <b>A2-whistle</b>  | <A2>be.able              | Q         |
| ‘Can you whistle?’ |                          |           |

The most efficient way to account for this categorization issue is to assume that *zudé* can be used either as a sound root with the usual associated features, or as an iconic verb integrated into the grammar, like *héch<sup>hi</sup>n* ‘to sneeze’ presented in Section 2.2.3. The difference between *zudé* ‘to whistle’ and *héch<sup>hi</sup>n* ‘to sneeze’ is that the first is converted from a sound root.

### 5.3 Semantic extensions

Among the numerous derivations of sound roots with instrumental prefixes, there are a few examples of semantic extensions from sound to visual effects, textures, or

movements. According to Dorsey (1892), the root °*dáze* refers to the sound of thunder. But *nádadaze* is defined as ‘to send out light in streamers or fan-like rays’ in Dorsey’s dictionary, ‘to send out sparks’ in his texts, and ‘to sparkle’ in Saunsoci & Eschenberg (2016). The latter specify that “[this verb] refers mostly to jewelry”, as illustrated in (41).

(41) (Saunsoci & Eschenberg 2016: 181 / speaker: Alice Saunsoci)

*Wanó<sup>n</sup>pi<sup>i</sup> k<sup>h</sup>e ná-dadaza.*  
 necklace the:HORIZ INS:temp-SOUND.REDUPL  
 ‘The necklace sparkled.’

A similar phenomenon occurs with the bound root °*p<sup>h</sup>úki* ‘popping sound’ combined with the derivational prefix *ná-* ‘by extreme temperature’. The verb *náp<sup>h</sup>uki* is defined as ‘to be made soft and light by the action of yeast, as bread’ (Dorsey n.d.). While this verb originally probably referred to the sound of rising dough, Dorsey makes reference to the resulting texture, and not the sound heard in the process.

Dorsey (1892) introduces the root °*shtáki* for ‘flapping or slapping sound, made in mud or some other soft object’. He also documents in his dictionary the root °*stáki* which means ‘flying off, as drops of water flung, or as a chip of wood that is hit with an ax’.<sup>28</sup> This could correspond to a semantic extension from sound to movement, along with a shift in the consonant grade.

Semantic extensions from sound to other sensory perceptions (such as movement, visual patterns, and textures) are well described for ideophone systems (see in particular McLean 2021).

## 6. Conclusion

This paper studies the numerous sound-denoting words of the Umó<sup>n</sup>ho<sup>n</sup> language, focusing in particular on one lexical class that I call “sound root”, in order to determine its relation to the cross-linguistic concept of ideophones.

---

<sup>28</sup> Dorsey writes it with an aspirated /k<sup>h</sup>/, but Rankin (1974) notes that Dorsey does not record aspiration consistently: the plosives that he writes as non-aspirated are never aspirated, while the plosives he writes as aspirated can be aspirated or non-aspirated.

Root	Completely depictive	- Grammatical integration +			Completely descriptive
<i>xwí</i>	<b>onomatopoeia</b> <i>xwííí</i> 'sounds of tree falling' (DD, DT)				
<i>k<sup>h</sup>uuu</i>	<b>onomatopoeia</b> <i>k<sup>h</sup>uuu</i> , <i>k<sup>h</sup>u</i> 'bang' (see 6)	<b>impersonal predicate</b> <i>k<sup>h</sup>ú</i> 'there is a <i>ku</i> sound made by shooting' (see 7)			
<i>hu</i>	<b>onomatopoeia</b> <i>hu-hu-hu</i> 'barking sound' (D92)	<b>verb modifier</b> <i>hu</i> , <i>úhuhu</i> 'barking' (see 38)	<b>intransitive verb</b> <i>hu</i> , <i>úhuhu</i> 'to bark' (see 39)	<b>transitive verb</b> <i>úhuhu</i> 'to bark at him/her' (DD, RT)	<b>noun</b> <i>hú</i> 'voice', 'sound' (DD, DT, RT, ULCC)
<i>˚táxi</i>		<b>used with égo<sup>n</sup></b> <i>mutáxi égo<sup>n</sup></i> 'to make the sound heard by firing a gun (...)' (see 25)	<b>verb with t<sup>h</sup>ígt<sup>h</sup>e</b> <i>taxi thatáxit<sup>h</sup>ígt<sup>h</sup>e</i> 'crunching of bones' (proper name, FLF)	<b>intransitive verb</b> <i>no<sup>n</sup>tátaxi</i> 'to make the sound at every step' (DT)	<b>transitive verb</b> <i>gatáxi</i> 'to make it give a tapping sound by hitting it or throwing it' (DT)
<i>bthó<sup>n</sup>xe</i>		<b>verb(?) with t<sup>h</sup>ígt<sup>h</sup>e</b> <i>bthó<sup>n</sup>xe t<sup>h</sup>ígt<sup>h</sup>e</i> 'to make a sudden crunching sound' (DD)	<b>intransitive verb (inanimate subject)</b> <i>bthó<sup>n</sup>btho<sup>n</sup>xe</i> 'to snap, as ice when forming' (DD)	<b>intransitive (?) verb (animate subject)</b> <i>thabthó<sup>n</sup>xe</i> 'to make a crunching sound once by gnawing' (DD)	<b>transitive verb</b> <i>babthó<sup>n</sup>xe</i> 'to make ice, etc., give a crunching sound, by pushing or punching at it' (DD)
<i>sathú</i>		<b>verb modifier / intr. verb</b> <i>sathú</i> 'rattling/ to rattle' (DT)	<b>intransitive verb</b> <i>thasáthu</i> 'to make the sound <i>sathú</i> by eating' (see 29)	<b>labile verb</b> <i>thisáthu</i> 'to rattle (it)' (see 37)	<b>noun</b> <i>sathú</i> 'rattlesnake' (FLF, ST, SLW)
<i>zúde</i>		<b>used with égo<sup>n</sup></b> <i>zudégo<sup>n</sup></i> 'with a whirr (of the wings)' (DD)	<b>verb modifier</b> <i>zuzúde wa'ó<sup>n</sup></i> 'to whistle, as a man does' (DD)	<b>intransitive verb</b> <i>bizúde</i> 'to wheeze, as when the nasal passages are obstructed' (DD)	
			<b>intransitive verb</b> <i>zúde</i> 'to whistle' (see 40)	<b>transitive verb</b> <i>zuzúde</i> 'to deceive a person in sport by averting the head after whistling to attract his attention' (DD)	
<i>˚ch<sup>h</sup>ízhe</i>			<b>intransitive verb</b> <i>gach<sup>h</sup>ízhe</i> 'to fall with a crash' (FLF)	<b>transitive verb</b> <i>bach<sup>h</sup>ízhe</i> 'to make a single cracking sound by pushing against a twig or small branch, which is broken by the act' (D92)	

Table 2: Continuum from less to more grammatically integrated sound-denoting root

I have defined sound roots on syntactic and semantic grounds: they are verbal roots that can combine with instrumental prefixes and whose meaning refers to sound. In Sections 3.3 through 3.6, I presented some morphological and syntactic features commonly attested with sound roots, setting them aside from the rest of the lexicon. However, these features are not attested with all sound roots, a fact possibly explained by the impoverished data.

In the current state of documentation, sound roots do not fit Dingemanse's (2019) comparative concept of ideophones, because they are not necessarily words (many are bound roots), they are never only depictive, and they are not systematically characterized by the features described in Sections 3.3–3.6. However, they show a resemblance to ideophones in other languages. In Section 5.3, I showed that semantic extensions are attested in Umó<sup>n</sup>ho<sup>n</sup> from sound to movement or texture. I also showed that the sound roots show different degrees of grammatical integration.

In Section 5.2, I showed how sound roots extend on a continuum from mostly depictive lexemes to completely descriptive lexemes. Onomatopoeia and sound roots are lexical classes that cover distinct areas in this continuum. At one extreme we find onomatopoeia: they are never syntactically integrated, and they do not undergo derivation to other parts of speech. By contrast, sound roots are never fully depictive, but by being sound-denoting, they retain some depictive force. They are attested in an array of morphological and syntactic environments that can broadly be ordered from less to more grammatically integrated.

## **Acknowledgments**

This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the program “Investissements d’Avenir” (reference: ANR-10-LABX-0083). It contributes to the IdEx Université Paris Cité – ANR-18-IDEX-0001.

I am deeply grateful to Aimée Lahaussais, Yvonne Treis, Armik Mirzayan and one anonymous reviewer for their insightful comments on previous versions of this paper. I am also grateful to all the Umó<sup>n</sup>ho<sup>n</sup> participants to the language workshops I organized in May 2023 to elicit sound roots.



## Abbreviations

1 = 1 <sup>st</sup> person	INTENS = intensifier	RT = Rudin et al. (1989-92)
2 = 2 <sup>nd</sup> person	M = masculine	SE = Saunsoci & Eschenberg (2016)
3 = 3 <sup>rd</sup> person	NEU = neutral (semantically bleached)	SG = singular
A = agentive person marker	OBV = obviative	SIT = sitting
AUX = auxiliary	onom = onomatopoeia	SLW = Sanchez, Larson & Walker (in progress)
conj. = conjunction	P = patientive person marker	SOUND = sound imitation
D92 = Dorsey 1892	PL = plural	ST = Stabler & Swetland (1977, 1991)
DD = Dorsey (no date)	PP = proximate/plural	temp = temperature
DECL = declarative	PX = proximate	ULCC = ULCC (2018)
DEM = demonstrative	Q = question marker	UNL = UNL-OLIT (2018)
DT = Dorsey (1890, 1891)	REDUPL = reduplication	VERT = vertical
EVID = evidential marker	REFL = reflexive	VOC = vocative
FLF = Fletcher & La Flesche (1911)	REPORT = reportative (subtype of evidential marker)	° indicates a bound root
HAB = habitual marker	RND = round	
HORIZ = horizontal		
IMP = imperative		
INS = instrumental prefix		

## References

- de Reuse, Willem J. 2022. Consonant symbolism in Western Apache ideophones. (Paper presented at the workshop Consonant and Vowel Symbolism in Native North American languages, Paris, December 8, 2023)
- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6(10). 654–672.
- Dingemanse, Mark. 2017. Expressiveness and system integration: On the typology of ideophones, with special reference to Siwu. *Language Typology and Universals* 70(2). 363–384.
- Dingemanse, Mark. 2019. ‘Ideophone’ as a comparative concept. In Akita, Kimi & Pardeshi, Prashant (eds.), *Ideophones, mimetics and expressives*, 13–34. Amsterdam: John Benjamins.
- Dingemanse, Mark & Akita, Kimi. 2016. An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics* 53(3). 501–532.

- Dorsey, James Owen. 1890. *The Čegiha Language*, Contributions to North American Ethnology, vol. 6. Washington, D. C.: U.S. government.
- Dorsey, James Owen. 1891. *Omaha and Ponka letters*. Bulletin 11. Washington: Bureau of American Ethnology.
- Dorsey, James Owen. 1892. Siouan onomatopes. *American Anthropologist* 5(1). 1–8.
- Dorsey, James Owen. n.d. A grammar and dictionary of the Ponca language by the Reverend J. Owen Dorsey - Formerly missionary under the Indian Commission of the Protestant Episcopal Church. (Manuscript dictionary flashcards)
- Duncan, Philip. T. 2022. How Kiksht does a lot with ideophones. (Paper presented at the workshop Typology of Ideophones, online, June 24-25, 2022)
- Dwyer, David & Moshi, Lioba. 2003. Primary and grammaticalized ideophones. In Mugane, John M. (ed.), *Linguistic typology and representation of African languages*, 173–185. Trenton: Africa World Press.
- Eschenberg, Ardis. 2005. *The article system of Umó<sup>n</sup>ho<sup>n</sup> (Omaha)*. Doctoral dissertation, State University of New York, Buffalo.
- Fletcher, Alice & La Flesche, Francis. 1911. *The Omaha tribe*. Washington: Bureau of American Ethnology.
- Guillaume, Antoine. 2009. Les suffixes verbaux de mouvement associé en cavineña. *Faits de Langues* 34(2). 181–204.
- Güldemann, Tom. 2008. *Quotative indexes in African languages*. Berlin: De Gruyter Mouton.
- Headman, Louis & O’Neil, Sean. 2019. *Dictionary of the Ponca People*. Lincoln: University of Nebraska Press.
- Heine, Bernd & Kuteva, Tania. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Jacques, Guillaume. 2023. Periodic tense markers in the world’s languages and their sources. *Folia Linguistica* 57(3). 539–562. (doi: <https://doi.org/10.1515/flin-2023-2013>)
- Johansson, Niklas Erben & Anikin, Andrey & Carling, Gerd & Holmer, Arthur. 2020. The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology* 24(2). 253–310.
- Körtvélyessy, Lívía. 2020. Onomatopoeia—A unique species? *Studia Linguistica*. 74(2). 506–551.
- Larson, Rory. 2022. RE: Could there be a suffix -é? (Email sent on June 16, 2022 to siouan@listserv.unl.edu)

- Marsault, Julie. 2021. *Valency-changing operations in Umó<sup>n</sup>ho<sup>n</sup>: affixation, incorporation, and syntactic constructions*. Doctoral dissertation, Université Sorbonne Nouvelle - Paris 3.
- Marsault, Julie. accepted. Consonant gradation in Umó<sup>n</sup>ho<sup>n</sup> and Pá<sup>n</sup>ka. Accepted in *Anthropological Linguistics*.
- Matthews, G. H. 1970. Some notes on the Proto-Siouan continuants. *International Journal of American Linguistics* 36(2). 98–109.
- McLean, Bonnie. 2021. Revising an implicational hierarchy for the meanings of ideophones, with special reference to Japonic. *Linguistic Typology* 25(3). 507–549.
- Merrill, John Thomas. 2018. *The historical origin of consonant mutation in the Atlantic languages*. Doctoral dissertation, University of California Berkeley.
- Mithun, Marianne. 1999. *The languages of Native North America*. Cambridge: Cambridge University Press.
- Mithun, Marianne. 2015. Morphological complexity and language contact in languages indigenous to North America. *Linguistic Discovery* 13(2). 37–55.
- Mithun, Marianne. 2017. Polysynthesis in North America. In Fortescue, Michael & Mithun, Marianne & Evans, Nicholas (eds.), *The Oxford handbook of polysynthesis*, 235–259. Oxford: Oxford University Press.
- Munro, Pamela. 1998. Chickasaw expressive “say” constructions. In Hinton, Leanne & Munro, Pamela (eds.), *Studies in American Indian languages: Description and theory*, 180–86. Berkeley: University of California Press.
- Rankin, Robert L. 1974. Observations on Dhegiha (Siouan) phonetics and phonology. (Paper presented at the XIIIth conference on American Indian Languages, November 22th 1974, Mexico D.F.)
- Rankin, Robert L. 2008. James Owen Dorsey’s The Dhegiha language. Unpublished MS.
- Rankin, Robert L. & Carter, Richard T. & Jones, A. Wesley & Koontz, John E. & Rood, David S. & Hartmann, Iren. (eds.). 2015. *Comparative Siouan Dictionary*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at: <https://csd.cld.org/> (Accessed 2023.11.20)
- Sanchez, Emmaline Walker & Larson, Rory & Walker, Arlene. in progress. *A phonological dictionary of Omaha, with Emmaline Walker Sanchez*. Unpublished MS.
- Saunsoci, Alice & Eschenberg, Ardis. 2016. *500+ verbs in Umó<sup>n</sup>ho<sup>n</sup> (Omaha) - doing things in the Umó<sup>n</sup>ho<sup>n</sup> way*. Scotts Valley, CA: CreateSpace Independent Publishing Platform.

- ULCC. 2018. Part I: Lessons from the Umó<sup>n</sup>ho<sup>n</sup> Language and Culture Center, Umó<sup>n</sup>ho<sup>n</sup> Nation Public School, Macy, Nebraska. *Umó<sup>n</sup>ho<sup>n</sup> Íye t<sup>h</sup>e, Umó<sup>n</sup>ho<sup>n</sup> Úshko<sup>n</sup>-t<sup>h</sup>e: The Omaha language and the Omaha way. An introduction to Omaha language and culture*, 3–242. Lincoln: University of Nebraska Press.
- UNL-OLIT. 2018. Part II: Lessons from the Omaha language class at the University of Nebraska-Lincoln. In ULCC & OLIT (eds.), *Umó<sup>n</sup>ho<sup>n</sup> Íye t<sup>h</sup>e, Umó<sup>n</sup>ho<sup>n</sup> Úshko<sup>n</sup>-t<sup>h</sup>e: The Omaha language and the Omaha way. An introduction to Omaha language and culture*, 243–636. Lincoln: University of Nebraska Press.
- Ullmann, Stephen. 1962. *Semantics: An introduction to the science of meaning*. Oxford: Basil Blackwell.
- Wilkins, David P. 1992. Interjections as deictics. *Journal of Pragmatics* 18(2–3). 119–158.

**CONTACT**

julie.marsault@inalco.fr

## Appendix A

The following table reproduces all the Umó<sup>n</sup>ho<sup>n</sup> material provided in Dorsey (1892), ordered alphabetically by roots (see Section 1.2). The definitions are unaltered from Dorsey, italics are preserved whenever Dorsey uses them, and the “etc.” following some examples are also quotes from the original.

The first column numbers the roots from 1 to 53, in bold. The second column numbers the entries in gray. Entries are roots, lexemes or examples associated with a definition or translation. A few roots have no definition, in which case Dorsey directly provides derived forms, like the root n°4 *ch<sup>h</sup>ízhē*. Since they have no definition nor translations, they are not numbered as entries, but they are numbered as roots.

The third column lists the roots, lexemes, or illustrative examples. The roots are in bold. Stress marks are reproduced as indicated by Dorsey (sometimes there is no stress).

The fourth column lists the definition or translation of each root, lexeme, or example. Italics are used wherever Dorsey uses them. Additions to the original occur in square brackets.

The roots and their derived forms that are excluded from the present study are in grayed cells (see Section 2.2).

	<b>Root</b>	<b>Definition</b>
	<b>lexeme; example</b>	
<b>1</b>	1 <i>’áxe</i>	used to describe the sounds of filing, grating, gnawing, or scratching on metal, bone, hard wood, etc.
	2 <i>no<sup>n</sup>’áxe</i>	the sound made by a horse when walking on frozen grounds
<b>2</b>	3 <i>bthó<sup>n</sup>’xe</i>	crunching sound, such as is heard when one eats a crust of bread or when a horse eats oats or corn, a dog gnaws a bone and crushes it, or as when one crushes through ice or snow
<b>3</b>	4 <i>ch<sup>h</sup>’áki</i>	[the sound heard in <i>gach<sup>h</sup>’áki</i> ]
	5 <i>gach<sup>h</sup>’áki</i>	to make the sound heard in clapping the palms of the hands together
<b>4</b>	<i>ch<sup>h</sup>’ízhē</i>	—
	6 <i>ch<sup>h</sup>’ich<sup>h</sup>’ízhē</i>	denotes the frequent crackling or breaking of twigs and small branches, or the frequent discharge of fire-arms

	<b>Root lexeme; example</b>	<b>Definition</b>
7	<i>bach<sup>h</sup>ʔzhe</i>	1. to make a single cracking sound by pushing against a twig or small branch, which is broken by the act 2. to push ahead, as through a thicket
8	<i>ákipach<sup>h</sup>ʔzhe</i>	to persevere in a certain course of conduct, despite all obstacles, regardless of the consequences
5	<i>dáʔe</i>	—
9	<i>no<sup>n</sup>dáʔe</i>	refer[s] to the sound of a horse's feet on hard, but not frozen, ground
10	<i>thidáʔe</i>	refers to one of the sounds of thunder, ʔʔʔ, whence we have the personal name <i>Wathídáʔe</i>
11	<i>Wathídáʔe</i>	Thunder being makes the sound ʔʔʔ ! (proper name)
12	<i>hú t<sup>h</sup>e dáʔe</i>	the voice is hoarse
6	<i>dáʔe</i>	—
13	<i>thidáʔe</i>	refers to the sound of the thunder, zzz, whence the personal name, <i>Wathídáʔe</i>
14	<i>Wathídáʔe</i>	Thunder being makes the sound zzz!
7	15 <i>dázhe</i>	all verbs in <i>dazhe</i> except one, <i>mudazhe</i> (...), refer to chafing or blistering the hands or feet
	16 <i>mudazhe</i>	<i>to fillip with the fingers</i>
8	<i>dúʔe</i>	— (“ <i>Dúʔe</i> has several derivatives”)
17	<i>thadúʔe</i>	to make the sound heard when a hazelnut is cracked between the teeth
18	<i>thidúʔe</i>	to make the sound heard when a stick is broken in the hands
9	<i>dúzhe</i>	—
19	<i>thidúzhe</i>	to split or crack a board by boring; to crack an egg by handling
10	20 <i>gíʔe</i>	the creaking of new shoes and the sound of fiddle-strings ( <i>Gi-gi-gi</i> )
	21 <i>gi-gi-gi</i>	the creaking of new shoes and the sound of fiddle-strings
	22 <i>bagíʔe</i>	to play the fiddle (i.e., make it creak by pushing the bow)
	23 <i>no<sup>n</sup>gíʔe</i>	to make (shoes) creak in walking
	24 <i>thagíʔe</i>	to gnash the teeth
11	25 <i>hák<sup>h</sup>ugthe</i>	whippoorwill

	<b>Root lexeme; example</b>	<b>Definition</b>
	<i>hák<sup>h</sup>ukthe!</i> 26 <i>hák<sup>h</sup>ukthe!</i> <i>azhó<sup>n</sup>.</i>	The Ponka children give the cry of the whippoorwill as follows: <i>hák<sup>h</sup>ukthe! hák<sup>h</sup>ukthe! azhó<sup>n</sup>.</i>
12	27 <i>héch<sup>h</sup>i<sup>n</sup></i>	<i>to sneeze</i>
13	28 <i>hu</i>	<i>to bark as a dog or wolf</i> is explained by the Omaha description of a barking sound: <i>hu-hu-hu!</i>
	29 <i>hu-hu-hu!</i>	barking sound
14	30 <i>húxpe</i>	<i>to cough</i>
15	31 <i>k'áxe</i>	used to describe the sounds of filing, grating, gnawing, or scratching on metal, bone, hard wood, etc.
	32 <i>thak'áxe</i>	mark[s] the sound made by rats when gnawing
16	33 <i>k'éxe</i>	used to describe the sounds of filing, grating, gnawing, or scratching on metal, bone, hard wood, etc.
	34 <i>thak'éxe</i>	mark[s] the sound made by rats when gnawing
17	35 <i>k'ushi</i>	denotes a gulping sound
	36 <i>thak'úshi</i>	to make noise by drinking
	37 <i>thak'úshi-xti</i> <i>nazhí<sup>n</sup></i>	said of the noise made in drinking water, whether by a horse or a person [- <i>xti</i> : intensifier; <i>nazhí<sup>n</sup></i> : to stand]
18	38 <i>kámo<sup>n</sup></i>	refers to the ringing of a bell, etc.
	39 <i>thikámo<sup>n</sup></i>	to ring a bell by pulling a rope
	40 <i>gakámo<sup>n</sup></i>	to strike, as a clock does
19	41 <i>kátho<sup>n</sup></i>	denotes the sound made in pushing against, or pulling from, a door, plank, or hard buffalo hide
	42 <i>bakátho<sup>n</sup></i>	denoting the action by pushing
	43 <i>thikátho<sup>n</sup></i>	... by pulling
	44 <i>gakátho<sup>n</sup></i>	... by hitting
	<i>etc.</i>	
20	45 <i>káxe</i>	a crow

	Root lexeme; example	Definition
21	46 <i>k<sup>h</sup>uuu</i> <sup>29</sup>	describes the report of a gun
22	<i>ku</i>	—
	47 <i>gakú</i>	<i>to beat a drum</i>
23	48 <i>kúge</i>	another drumming sound [difficult to distinguish from <i>tíde</i> ]
24	- <i>mo<sup>n</sup></i>	—
	49 <i>bamó<sup>n</sup></i>	to use a file, sharpen a scythe by pushing (the original reference (...) may have been to the sound made)
	50 <i>thimó<sup>n</sup></i>	to sharpen an ax on a grindstone (the original reference (...) may have been to the sound made)
	51 <i>bimó<sup>n</sup></i>	to knead dough
25	52 <i>p<sup>h</sup>úki</i>	a popping sound, as heard in drawing a cork from a bottle, or a deadened sound, a sort of thud, as in hitting flesh, garments, or other soft objects
	53 <i>gap<sup>h</sup>úkitho<sup>n</sup></i> <i>gáxe</i>	to make the sound heard when one lets a book fall to the floor or ground
26	54 <i>s'u</i>	resembles the sound heard in planing ( <i>s! s! s!</i> )
	55 <i>bas'ú</i>	<i>to plane</i>
	56 <i>this'ú</i>	<i>to use a drawing knife</i>
27	57 <i>sáp<sup>h</sup>i</i>	describes such a cracking or smacking sound as is made by a whip-lash
	58 <i>gasáp<sup>h</sup>i</i>	to use a whip
	59 <i>wégasáp<sup>h</sup>i</i>	a whip
	60 <i>gasáp<sup>h</sup>iitho<sup>n</sup></i> <i>gáxe</i>	to make the sound heard when one lets a book fall to the floor or ground
28	61 <i>sathú</i>	used in speaking of the rattling of corn in a granary or on a pile out of doors, as well as of the rattling of the <i>wes'a sathu</i> or rattle snake
	62 <i>wes'a sathu</i>	rattle snake
29	63 <i>shathú</i>	conveys two ideas, [1.] a swishing sound, made in water ; [2.] the sound made by the hitting, dragging, etc., of a chain.

<sup>29</sup> This root is written *k<u+* in Dorsey (1892). The sign + refers to expressive lengthening (written with a triplication of the vowel in this paper), but Dorsey does not specify what '<' means, and this character is not found in his other publications. It is rendered as an aspirated /k<sup>h</sup>/ here, based on the various attested examples of the onomatopoeia *k<sup>h</sup>uuu* in texts.



	Root lexeme; example	Definition
30	<i>shka</i>	–
	64 <i>múshkashka</i>	to gargle the throat
31	<i>shkáp<sup>hi</sup></i>	–
	65 <i>gashkáp<sup>hi</sup></i>	to make the sound heard in slapping the cheek of the back of the hand
32	66 <i>shtáki</i>	describe[s] one (...) flapping or slapping sound, made in mud or some other soft object
	67 <i>shtashtáki</i>	describe[s] [more than one] flapping or slapping sounds, made in mud or some other soft object
33	68 <i>skáp<sup>hi</sup></i>	[the sound heard in <i>gaskáp<sup>hi</sup></i> ]
	69 <i>gaskáp<sup>hi</sup></i>	to make the sound heard in clapping the palms of the hands together
34	70 <i>sú'e</i>	applied to two sounds: (1.) <i>sss</i> , the sound of ice breaking up and floating off, or that of a steady rain; (2.) <i>sk! sk! sk!</i> the swishing sound made in walking through grass
35	71 <i>támo<sup>n</sup></i>	refers to the ringing of a bell, etc.
	72 <i>thitámo<sup>n</sup></i>	to ring a bell by pulling a rope
36	73 <i>táshi</i>	seems to be used in two ways [see derived forms]
	74 <i>batáshi</i>	to make the sound heard when one taps on a table with the end of a pencil
	75 <i>gatáshi</i>	to make the sound <i>tshshsh</i> heard when one strikes a tree with an ax when the sap is flowing
37	76 <i>tási</i>	refers to a snapping sound, made by the aid of a rope, cord, or stiff hide
	77 <i>thitási égo<sup>n</sup></i>	to make such a sound by pulling a cord
	78 <i>batási</i>	to make a snapping sound by punching against a rope or stiff hide
38	79 <i>táxi</i>	is formed from the sound <i>txxx</i>
	80 <i>gatáxi</i>	to make the sound heard when a tree is struck with an ax in cold weather
	81 <i>thitáxi égo<sup>n</sup></i>	describes a sound of thunder, <i>t'kh-t'kh-hyuuu!</i>
	82 <i>t'kh-t'kh-hyuuu!</i>	a sound of thunder
	83 <i>batáxi</i>	is used of the sound heard when one pushes suddenly against a bone

	Root lexeme; example	Definition
	84 <i>no<sup>n</sup>táxi</i>	refers to the sound of a horse's feet on hard, but not frozen, ground
39	85 <i>tíde</i>	refer[s] to a hollow or drumming sound on the floor, the ground, or a door (difficult to distinguish from <i>kúge</i> )
	86 <i>no<sup>n</sup>títide</i>	[make a pattering sound while walking]
40	87 <i>túp<sup>h</sup>i</i>	marks a pattering sound, as in <i>no<sup>n</sup>tútup<sup>h</sup>i</i>
	88 <i>no<sup>n</sup>tútup<sup>h</sup>i</i>	[to make a pattering sound while walking]
41	89 <i>túshi</i>	describes the crackling of twigs, the report of a gun, etc.
	90 <i>batúshi</i>	to fire a popgun – i.e., by pushing
	91 <i>thitúshi</i>	to snap the fingers, to fire a gun once – i.e., by pulling the trigger
	92 <i>tutúshi</i>	denotes the frequent crackling or breaking of twigs and small branches, or the frequent discharge of fire-arms
42	<i>túxi</i>	–
	93 <i>thitúxi</i>	marks a crackling sound made by pulling
43	94 <i>xátho<sup>n</sup></i>	describes the sound made in brushing against or pulling through sun-flowers, grass, or leaves
	95 <i>baxátho<sup>n</sup></i>	denoting the action by pushing
	96 <i>thixátho<sup>n</sup></i>	... by pulling
	97 <i>gaxátho<sup>n</sup></i>	... by hitting
	<i>etc., etc</i>	
44	98 <i>xtházhe</i>	to scream or cry out, as a young animal does
	99 <i>thaxtháxtházhe</i>	to talk or sing in a quavering voice
	100 <i>bixtháxthazhe</i>	to make a flute give forth quavering notes, <i>etc., etc.</i>
45	101 <i>xthó<sup>n</sup>zhe</i>	marks a crunching sound
	102 <i>baxthó<sup>n</sup>zhe</i>	to crush an egg-shell by pushing at it
	103 <i>thixthó<sup>n</sup>zhe</i>	to make the crunching sound heard when a sled is pulled over firm snow on a frosty morning
46	<i>xthúde</i>	–
	104 <i>zho<sup>n</sup>xthúde</i>	<i>to snore</i>

	<b>Root lexeme; example</b>	<b>Definition</b>
47	105 <i>xú'e</i>	the sound of which is given as <i>xxx</i> , describes the sound made in tearing calico, the roar of falling water (whence, <i>ni xu'e</i> , a <i>waterfall</i> ), the sound heard in sawing or in scraping wood by pushing, as well as the whizzing of a whirled stick
	106 <i>ni xu'e</i>	<i>a waterfall</i>
48	<i>za</i>	– [Dorsey proposes a link with <i>záde</i> ]
	107 <i>no<sup>n</sup>bé ugáza</i>	<i>the phalanges</i> (noun)
49	108 <i>za'e</i> <sup>30</sup>	a noise, buzz, confusion applied to the sound of millstones in motion
50	109 <i>záde</i>	usually conveys the idea of branching off or forking...
	110 <i>hú t<sup>h</sup>e záde</i> <i>ínahi<sup>n</sup></i>	the voice is really indistinct – that is, the sound scatters instead of going straight to the person addressed
51	111 <i>zhudé</i>	refers to the expulsion of the breath by a person or animal that is nearly exhausted from running, etc.
52	112 <i>zúde</i>	denot[es] a hissing sound of confined air that is escaping
	113 <i>názide, nájije</i> <sup>31</sup>	to make a sizzling sound, as when meat is broiling
53	114 <i>zúde</i>	denotes a whistling sound, such as a man makes
	115 <i>zuzúde wa'ó<sup>n</sup></i>	to whistle a tune, as a man does
	116 <i>gazúzude</i>	to roar or whistle often, as the wind does

<sup>30</sup> Dorsey does not write stress on this root in his paper, but it is attested as *za'é* in his texts.

<sup>31</sup> *Nájije* corresponds *názide* with expressive palatalization (with diminutive meaning).

## Appendix B

This table lists all the roots from 1 to 53 and specifies the part of speech they belong to and the features associated with each of them. The features are presented from Section 3.3 to Section 4.1, and summarized in Section 5.1.1. The roots excluded from the study are moved to the end of the table, after the line ‘Total’.

- **Part of speech of the root**

Sound roots are all verbs (v). When they seem to be bound roots, ‘bd’ is added in parentheses. It is not clear whether the roots only found in collocation with *t<sup>h</sup>ígthe* or *t<sup>h</sup>ítthe* are bound roots, so a question mark is added. The root *k’úshi* is used as a verb by at least one speaker according to Dorsey (1892: 3), while others only use it with derivational prefix (i.e., as a bound root). The roots eliminated from the data set in Section 2.2. include various parts of speech, including ‘n’ for ‘noun’ and ‘onom’ for ‘onomatopoeia’.

- **Suffix -é**

‘Y’ means that the root (with or without an instrumental prefix) is attested with the suffix -é.

- **Collocation with *t<sup>h</sup>ígthe* or *t<sup>h</sup>ítthe***

‘Y’ means that the root (with or without an instrumental prefix) is attested followed by *t<sup>h</sup>ígthe* or *t<sup>h</sup>ítthe*.

- **Collocation with *égo<sup>n</sup>***

Cells of this column are filled when a collocation with *égo<sup>n</sup>* ‘to be (somewhat) like it’ is attested for the root. I specify whether the construction with *égo<sup>n</sup>* concerns the root by itself (‘root’), the reduplicate root (‘RED’), the root with a prefix *wa-* or an instrumental prefix (‘ins.’).

- **Instrumental verbs with unusual features**

Cells of this column are filled when the root is attested in an instrumental verb with unusual features. The content of each cell specifies what type(s) of unusual feature(s):

- ‘intr.’ means the verb is intransitive, while a causative derivation is expected;
- ‘égo<sup>n</sup>’ means that it is attested in collocation with *égo<sup>n</sup>* ‘to be (somewhat) like it’. (Note that in this case, it is redundant with the preceding column.)
- ‘modif.’ means it is used as a verb modifier, as in example (28).
- ‘redun.’ means the prefix’s meaning is redundant with another verb, as in examples (28) to (31).

Numbers in parentheses specify how many examples were found for each type of feature. For instance, the root *kúge* is found in one example in an instrumental verb acting as a verb modifier, and where the prefix is semantically redundant. The root *sathú* is attested three times in intransitive instrumental verbs, one of which is also followed by *égo<sup>n</sup>*.

- **Consonant gradation (CG)**

‘Y’ indicates that the root is linked to another by consonant gradation; ‘—’ indicates that the root has no fricative; an empty cell means that the root is not linked to another by consonant gradation, even though it has a fricative.

Root	Part of speech	-é	<i>t<sup>h</sup>ígthe,</i> <i>t<sup>h</sup>ítthe</i>	<i>égo<sup>n</sup></i>	Instrumental verbs with unusual features	CG
1	<i>’áxe</i>	v		Y		
2	<i>bthó<sup>n</sup>xe</i>	v (bd?)		Y		Y
3	<i>ch<sup>h</sup>áki</i>	v (bd)				—
4	<i>ch<sup>h</sup>ízhe</i>	v (bd)				Y
5	<i>dáxe</i>	v (bd)		ins.	intr. & <i>égo<sup>n</sup></i> (1)	Y
6	<i>dáze</i>	v (bd)				Y
7	<i>dúxe</i>	v (bd)				Y
8	<i>gíze</i>	v				Y
9	<i>k’áxe</i>	v (bd)			intr. (1)	
10	<i>k’éxe</i>	v (bd)				
11	<i>k’ushi</i>	v (bd?)	Y	Y	intr. (1)	

	Root	Part of speech	-é	<i>t<sup>h</sup>ígthe,</i> <i>t<sup>h</sup>ítthe</i>	<i>égo<sup>n</sup></i>	Instrumental verbs with unusual features	CG
12	<i>kámo<sup>n</sup></i>	v (bd)					—
13	<i>kátho<sup>n</sup></i>	v (bd)				<i>t<sup>h</sup>ítthe</i> (1)	—
14	<i>ku</i>	v (bd)					—
15	<i>kúge</i>	v				redun. & modif. (1)	—
16	<i>p<sup>h</sup>úki</i>	v (bd)	Y	Y	ins.	<i>égo<sup>n</sup></i> (2); <i>t<sup>h</sup>ígthe</i> (1)	—
17	<i>s'ú</i>	v (bd)					
18	<i>sáp<sup>h</sup>i</i>	v			ins.	<i>égo<sup>n</sup></i> (1)	
19	<i>sathú</i>	v (bd?)			<i>wa-</i>	intr (2); redun. & intr. (2)	Y
20	<i>shathú</i>	v (bd)					Y
21	<i>shka</i>	v (bd)					
22	<i>shkáp<sup>h</sup>i</i>	v (bd?)		Y	ins.	intr. & <i>égo<sup>n</sup></i> & <i>t<sup>h</sup>ígthe</i> (1); intr. & <i>égo<sup>n</sup></i> (2)	Y
23	<i>shtáki</i>	v (bd)					Y
24	<i>skáp<sup>h</sup>i</i>	v (bd)					Y
25	<i>sú'e</i>	v					Y
26	<i>támo<sup>n</sup></i>	v (bd)					—
27	<i>táshi</i>	v (bd)	Y	Y	ins.	intr. & <i>égo<sup>n</sup></i> (2)	Y
28	<i>tási</i>	v (bd)	Y	Y	ins.	<i>égo<sup>n</sup></i> (4)	Y
29	<i>táxi</i>	v (bd)	Y	Y	ins.	intr (1); redun. & intr. (1); <i>égo<sup>n</sup></i> (4); intr. & <i>égo<sup>n</sup></i> (1); <i>t<sup>h</sup>ígthe</i> (1)	Y
30	<i>tíde</i>	v (bd)		Y	ins.	intr. (2); intr. & <i>égo<sup>n</sup></i> (1)	—
31	<i>túp<sup>h</sup>i</i>	v (bd)					—
32	<i>túshi</i>	v	Y	Y	ins.	intr. & <i>égo<sup>n</sup></i> (1)	Y
33	<i>túxi</i>	v (bd)			ins.	intr. & <i>égo<sup>n</sup></i> (1)	Y
34	<i>xátho<sup>n</sup></i>	v (bd)					
35	<i>xtházhe</i>	v		Y			

	Root	Part of speech	-é	<i>t<sup>h</sup>ígthe</i> , <i>t<sup>h</sup>ítthe</i>	<i>égo<sup>n</sup></i>	Instrumental verbs with unusual features	CG
36	<i>xthó<sup>n</sup>zhe</i>	v					
37	<i>xthúde</i>	v (bd)				intr. (1)	
38	<i>xú'e</i>	v		Y	ins.	redun. & intr. & <i>égo<sup>n</sup></i> (1); redun. & intr. (1)	Y
39	<i>za'é</i>	v		Y			
40	<i>zhudé</i>	v					Y
41	<i>zíde</i>	v (bd)					
42	<i>zúde</i>	v			root, RED		Y
	<b>TOTAL</b>		<b>6</b>	<b>13</b>	<b>13</b>	<b>17</b>	<b>20</b>
43	<i>dazhe</i>	v (bd)					Y
44	<i>dúzhe</i>	v (bd)					Y
45	<i>hák<sup>h</sup>ugthe</i>	n					—
46	<i>héch<sup>h</sup>i<sup>n</sup></i>	v					—
47	<i>hu</i>	onom / v					—
48	<i>húxpe</i>	v					—
49	<i>káxe</i>	n					—
50	<i>k<sup>h</sup>uuu</i>	onom					—
51	<i>mo<sup>n</sup></i>	v (bd)				<i>t<sup>h</sup>ígthe</i> (1)	—
52	<i>za</i>	v (bd)					Y?
53	<i>záde</i>	v					Y?

# Teko ideophones: description of a word class

FRANÇOISE ROSE

CNRS, DYNAMIQUE DU LANGAGE & UNIVERSITÉ LUMIÈRE LYON 2

Submitted: 30/12/2022 Revised version: 26/07/2022

Accepted: 02/10/2023 Published: 12/06/2024



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

## Abstract

The aim of this paper is to present a comprehensive description of the ideophones of Teko, a Tupi language spoken in French Guiana. This word class, previously only briefly described, is defined in this paper through a systematic comparison to nouns and verbs, at various levels: phonology, word structure, prosody, semantic, morphology, syntax and discourse use. In particular, various aspects of prosody are investigated, including syllabic structure, pitch, intensity and duration, and pauses. The qualitative and quantitative analyses are based on a lexical database of 177 ideophones, 420 occurrences in texts, and a subset of 101 tokens with audio-recording. Contrary to the common view on ideophones that postulates a rather marginal status of the latter, this paper shows that ideophones are in fact rather well integrated in the linguistic system of Teko. Yet they show regularities that require them to be considered a distinct word category.

**Keywords:** parts of speech; phonosymbolism; prosody; reduplication; pause; expressivity.

## 1. Introduction

The aim of this paper is to present a comprehensive description of the ideophones of Teko (Glottocode eme 1243). This word class was previously only briefly described (Rose 2011: 400-409) as a part-of-speech illustrating a situation through expressive means, in line with Doke (1935: 118) or Dingemans (2019). In the following example, *tapug* illustrates a ‘diving’ situation through expressive means. In this context, the main characters of the myth (two sisters) dive into the river to escape their brother and thus turn into mermaids.



- (1) *tapug*      *ze-kapirer = ne*      *?i-b*      *o-po-por*      *o-ho-oŋ*.  
 IDEO.diving RECP-behind = CONTR water-in      3-RED-jump      3-go-PL.S  
 ‘Splash, they dive in the water one after the other and go away.’ 04.035

This paper contributes to the growing literature on the cross-linguistic definition and description of ideophones. More precisely, it aims to provide new data and analyses for the questions of whether ideophones are marginal in the lexicon, whether they are extra-grammatical, and whether they form a specific word class or not. To do this, ideophones are systematically compared to nouns and verbs, as concerns their phonology, prosody, semantic, morphology, syntax and discourse. This corpus-based multi-level comparison makes it clear that we are dealing with a separate word class.

Two methodological aspects make this study innovative. First, the comparison of ideophones with both nouns and verbs will be as systematic as possible. When relevant, quantitative analyses with statistical tests will be used to support the comparison. The objective is to avoid cherry-picking in the choice of examples or features to be put forward. Pointing to specific items or features runs the risk of over-exoticizing ideophones, by placing their characteristics in the foreground. Second, this study makes room for prosody, interpreting the term broadly, i.e. not just intonation, but also stress, syllabic structure of words, harmony processes, and pauses. The prosody of ideophones is not very well investigated, but some publications broach this topic (Nuckolls 1996; Kita 1997; Dingemanse 2017; Akita 2021; see also Smoll 2014: 20, for a list of other studies covering the prosody of ideophones).

The importance of studying the prosody of ideophones has been highlighted by the work of Akita (2021). It distinguishes three strategies for formally marking depictive (iconic) signs, which include ideophones. The strategies depend on which part of the utterance is concerned. In this framework, ‘framing’ marks the boundary between the ideophone and the rest of the utterance, and typically includes pauses and quotative markers; ‘foregrounding’ focuses on the characteristics highlighting the ideophone itself, such as prominent prosody; and ‘backgrounding’ focuses on the characteristics of the rest of the utterance (such as low pitch on a quotative verb). The present study investigates both the framing and foregrounding strategies to mark ideophones in speech.<sup>1</sup>

---

<sup>1</sup> The prosodic analysis of the data had already been carried out when the work of Akita (2021) was published, hence the lack of attention to the backgrounding strategy in the present paper.

The remainder of this section focuses on presenting the Teko language and the data. The following sections will focus in turn on segmental phonology (Section 2), root structure (Section 3), word-level prosody (Section 4), morphology (Section 5), syntax (Section 6), discourse use (Section 7), and semantics (Section 8). The final section (Section 9) will discuss the extent to which Teko ideophones are regular or marginal within the language system, and how they form a word class, by comparing their characteristics with those of Teko nouns and verbs.

### **1.1. Teko**

The Teko community consists of about 400 people living in two areas in French Guiana: next to the Maroni river (on the border with Suriname) and at the Oyapock-Camopi confluence (on the border with Brazil). The community (formerly known as Emerillon) is the result of the aggregation of surviving members of different small ethnic groups, mainly of Tupí-Guaraní origin (Navet 1994). The Teko language is still actively being used and passed on to children as a native language but it must nevertheless be considered endangered given the small number of speakers and the ever-increasing intensity of contacts with French and Guianese Creole speakers.

Teko belongs to the Mawetí-Guaraní group (and more precisely its Tupí-Guaraní sub-group) of the Tupí stock (Rodrigues 1984; Rose 2023). Tupí languages are spoken throughout Brazil, in northern Argentina, Paraguay, Bolivia, and French Guiana. Ideophones, or sub-classes of word classes with sound symbolic features, have been mentioned for a large number of the Tupí languages, with a few rather detailed studies (Langdon 1994; Gabas 1999: 234-263) and a full book devoted to the issue in Awetí (Reiter 2011). In the latter work, a chapter is devoted to the comparison of ideophones and verb structures in the family (Reiter 2011: 495-575). Due to space limitations, this paper will focus on the description of Teko ideophones, without a systematic comparison to other Tupí languages.

The Teko language was first described at the turn of the millennium (Maurel 1998). A first reference grammar is available (Rose 2011), with ten pages devoted to ideophones. The present paper will build on it. Among the other published work on the language, the following papers deal with issues of interest for our present purposes: the distinction between nouns and verbs (Couchili et al. 2002), stress (Gordon & Rose 2006), reduplication (Rose 2005; Rose 2007), phonology (Rose

2008). A short introduction to relevant aspects of the language will be given in each section of this paper.

## 1.2. *The sources*

The dataset has been built from two main sources: my corpus, and an online dictionary.

The corpus was collected between 1999 and 2004 in French Guiana, mainly in the village of Camopi, but also in Cayenne and its surroundings. It is made up of wordlists, elicited data and texts. This work is mainly based on the text corpus. It is made up of thirty-eight short texts, with a total of 2000 utterances.<sup>2</sup> Twenty-five texts have been recorded in the field either by me between 1999 and 2003 or by Alexis Michaud in 1998. From these twenty-five recordings, only twenty-one recordings are available amounting about 100 minutes. These are transcribed in ELAN (2022). The other thirteen texts are written texts that had been previously published (Renault-Lescure et al. 1987; Maurel 1993; Maurel 2000; Association Solidarité Guyane 2000; Maurel 1991). All the texts are transcribed, translated into French and English, and annotated (with parts-of-speech and translation) at the morpheme level with Toolbox<sup>3</sup>. The Teko toolbox project (Rose 2018) also comprises a lexicon with 1465 entries, each with parts-of-speech information, and translation in French and English. An excerpt has been made public at the AILLA<sup>4</sup> and Ortolang<sup>5</sup> archives.

This corpus is complemented by a recent French-Teko dictionary published online by a group of speakers (Cachine et al. 2020). It has 2539 entries. Data from this source is followed by an asterisk when cited as an example or in a table in this paper.

This study aims to describe ideophones on the basis of the corpus and the online dictionary. It is therefore exploratory in that it draws from a limited set of data, and hardly builds on discussion with speakers (only a little elicitation on ideophones was carried out during fieldwork). The primary results of this work consist of two ideophone databases (Section 1.4).

---

<sup>2</sup> When cited as examples in this paper, sentences from the corpus are followed by the text number and the sentence number, separated by a dot, as in 04.035 for sentence #35 in text #4.

<sup>3</sup> <https://software.sil.org/toolbox/>

<sup>4</sup> <https://www.ailla.utexas.org/collections/549/>

<sup>5</sup> <https://hdl.handle.net/11403/sldr000870>

### 1.3. Presence of ideophones in the sources

Table 1 shows that the share of ideophones in the Teko lexicon varies from 5 to 8% depending on the source.<sup>6</sup> They represent 4% of the words in the text corpus, where they occur on average every five sentences. All in all, it is clearly not possible to consider this category as marginal in the Teko language and in speech.

	ideophones	nouns	verbs	roots/words	percentage of ideophones
dictionary	130	1155	819	2539	5%
Toolbox lexicon	108	541	270	1329	8%
Toolbox texts	420	2636	3133	10,068	4%

**Table 1:** Proportion of ideophones in the lexicon and texts.

### 1.4. Databases and datasets

Four databases accompany this paper as freely accessible supplementary materials in both csv and xlsx formats, in the *Teko ideophones* collection on the Ortolang platform.<sup>7</sup> They code for a large number of features. Most of these features, such as root structure, monovocality, reduplication, or types of syntactic integration, appeared important on the basis of observation of the Teko data since early data collection began, more than two decades ago (see Rose 2003). Note that some of these features happen to have been discussed in the subsequent general literature on ideophones, such as the unusual phonotactics in Yurakaré ideophones (Gijn 2010), reduplication in Japhug ideophones (Jacques 2013), monovocality in Japanese ideophones (Akita et al. 2013), and types of syntactic integration of Siwu ideophones (Dingemanse 2017). Other features coded in the databases are directly taken from the literature on ideophones in other languages, such as for example, deideophonic derivation (Reiter

<sup>6</sup> Affixes and clitics/particles have been excluded from the total number of entries in order to yield the number of lexical roots, and proper names have been excluded from the count of nouns, for both dictionary and Toolbox lexicon. Moreover, borrowings have been excluded from the count of nouns and verbs, for the Toolbox lexicon only. Nouns and verbs in the dictionary are overestimated because the roots with several sub-categories (such as intransitive verb and transitive verb) were counted several times.

<sup>7</sup> <https://hdl.handle.net/11403/teko-ideophones/v1>

2011: 325-334), or the semantic motion grid for ideophones developed for Basque ideophones (Ibarretxe-Antuñano 2019).

#### 1.4.1. *The Ideophone type database*

The *Ideophone type database* lists all known Teko ideophones up to date. For this, the 108 ideophones in my lexicon and the 130 ideophones from the dictionary (Cachine et al. 2020) have been merged and their orthography made consistent.<sup>8</sup> This yields a list of 177 ideophones. It is not meant to be definitive or comprehensive, but rather to reflect the present stage of analysis of potential ideophonic words.

These have been systematically coded for the features listed below, in the following order:

- Form, ID, and Phonological transcription
- Source, i.e. *Dictionnaire Teko* for Cachine et al. (2020) and/or *Toolbox Lexicon* for Rose (2018)
- Variants and derivation (3 features)
- Root structure
- Initial consonant, medial consonant(s), final consonant, vowel(s)
- Exclusively /r/ in medial position, Monovocality
- Syntactic constructions (4 features)
- Provisional gloss
- Semantic categorizations (4 features)
- Motion semantic grid (26 features)

---

<sup>8</sup> Ideophonic forms that were phonetically close and semantically related have been merged as one unit (see *pu* and *fu*; *tou*, *to* and *tuu*). Complex forms showing the repetition of the exact same sequence in what seems to be a multimorphemic word in the source have been treated as the repetition of a simpler ideophone (such as *tur* for *turtur*). What was obviously a reduplicated form of a more basic ideophone was not treated as a separate entry (such as *dururug* from the dictionary, compared to *durug* from the Toolbox lexicon). In all cases, the form taken as the reference form in this study is the simplest form of each ideophone to the best of my knowledge. Some doubts remain on the identification of some of the listed items as ideophones (*āhā* could be an interjection), on the relation between various items (is *tītītīg* a variant of *tīg* ?), and on what is the basic form of some ideophones (should we consider that *woroworoġ* is built on a putative basic form *woroġ* ?). These questions could be solved with access to more data or speaker's intuitions.

#### 1.4.2. *The Noun, Verb and Ideophone database*

In order to compare the form of the ideophones listed in the *Ideophone type database* with that of nouns and verbs, a dataset of nouns and verbs has been put together. The list was extracted from the Toolbox lexicon (Rose 2018), excluding proper nouns and borrowings. It consists of 541 nouns and 270 verbs. It is complemented by the list of ideophones, transcribed phonologically.

Each item has been coded manually for parts-of-speech, root structure, monovocality, /r/ in word-internal position.<sup>9</sup>

#### 1.4.3. *The Ideophone token database*

The *Ideophone token database* is a sample of 101 audio-recorded occurrences of ideophones extracted to investigate the use of ideophones in speech, especially their prosody.<sup>10</sup> The examples have been extracted from the text corpus, mainly narratives and one expository text. For each token, the *Ideophone token database* provides information about:

- ID, Form, Example number
- Reduplication and its meaning (2 features)
- Series of ideophones (3 features)
- Syntactic integration (5 features)
- Sentence type other than declarative
- Pause duration before and after ideophones

#### 1.4.4. *Vowel results*

The *Ideophone token database* database is complemented by a sheet called *Vowel results* giving the prosodic characteristics (duration, intensity and pitch) of the 183 vowels of the 101 ideophone tokens.

---

<sup>9</sup> These columns are not filled for ideophones in this database, as the information is already coded in the *Ideophone type database*.

<sup>10</sup> All ideophone tokens from the audio files of spontaneous speech were extracted, following the chronological order of the recordings, until a hundred tokens were reached. It happens that the very great majority of tokens have been extracted from a single text with a single speaker.

## 2. Phonology

This section compares the frequency and distribution of individual phonemes in ideophones, nouns and verbs. This comparison is based on the *Noun, Verb and Ideophone database* using the tool StatMe that allows an easy investigation of the frequency of phonemes in a list of linguistic items.<sup>11</sup>

### 2.1. Introduction to the phonology of Teko

The inventory of consonants and vowels is given in Table 2 and Table 3. Phonetic realizations are specified in brackets using the IPA notation.

	labial	alveolar	palatal	velar	labio-velar	glottal
voiceless non-continuant	p [p, p̚]	t [t, t̚]	tʃ [tʃ, tʃ̚]	k [k, k̚]	kʷ	ʔ
voiced non-continuant	b [b, ᵐb, m]	d [d, ᵐd, n]	ɕ [ɕ, ɲ]	g [g, ŋ]		
voiceless continuant		s [s, ts]				
voiced continuant		z [z, dz]				
non-obstruent	w [w, β, w̃]	r [r, ɻ, n]	j [j, j̃]			h [h, ɣ, h̃]

**Table 2:** Teko consonantal inventory.

	front	central	back
closed	i, ī	ɨ, ɨ̃	u, ũ
mid	e, ē [e, ε]	ə	o, õ [o, ɔ]
open		a, ã	

**Table 3:** Teko vowel inventory.

In word-initial position, all consonants but /g/, /j/ and /r/ are found. In word-internal position, all consonants are found, but /g/ is restricted to morpheme-final position. In word-final position, only non-continuants are found, and are then either unreleased [p̚], [t̚], [tʃ̚] and [k̚] or nasalized. Unreleased consonants have been analyzed as voiceless non-continuants, and nasal consonants as allophones of voiced non-continuants in Rose (2008).

<sup>11</sup> StatMe is a free access tool doing simple counts on the distribution of units in lexicon <https://reflex.cnrs.fr/STATme>.

Nasality in Teko results from regressive harmony within roots (and diffusion to adjacent affixes) affecting all voiced segments without being stopped by voiceless segments. Nasality diffuses from either nasal vowels as in /tiakã/ [tiãkã] ‘river’ or voiced non-continuants, which are phonetically realized nasal in word-final position as in /abad/ [ãmã] ‘rain’ and prenasal in intervocalic position as in /kadetat/ [kãndetat̃] ‘crown’.

In the writing system used in this paper (outside of phonemic and phonetic transcriptions), nasal consonants are written as such, nasal vowels are marked with tilde only when phonemically nasal, and final unreleased consonants are written as voiced consonants.<sup>12</sup>

## **2.2. Phonemic inventory in ideophones**

The phonemic inventory used in ideophones is overwhelmingly very similar to that used in the rest of the lexicon and grammatical elements. There are two sounds that are not part of the regular phonemic inventory and that are each found in only one (variant of) ideophone: [ʃ] in /tiʃ/ ‘grease dripping on fire’ and [f] in /fu ~ pu/ ‘blowing’.<sup>13</sup>

## **2.3. Phoneme frequency and distribution in ideophones**

Phonemes follow the same restrictions in distribution in ideophones as in the language in general. The consonants /g/, /j/ and /r/ are not found root-initially, /g/ is not found root-internally either. Only non-continuants fill the root-final position. Notable exceptions are the presence of /h/ word-finally in two ideophones only, as in /puh/\* ‘shaman blowing’, as well as one example with /ʃ/ (/tiʃ/, see 2.2).

The systematic comparison of ideophones with nouns and verbs nevertheless allows us to spot some differences in the frequency of attestations of phonemes across word classes.

All vowels are found in ideophones, with /u/ and /o/ being the most frequent (> 20%) and /e/ and /ə/ the least frequent ones (< 10%). Most notably, /o/ and /ə/

---

<sup>12</sup> All known writing systems note nasal consonants as such, and nasal vowels in a less systematic way. Final unreleased consonants are either written with symbols for voiceless consonants, as in the dictionary (Cachine et al. 2020), or with symbols for voiced consonants to ensure a unique root form whatever the morphophonological context, as in the grammar (Rose 2008).

<sup>13</sup> Both sounds are found in borrowings.



are much more frequent in ideophones than what is expected from their number in the overall lexicon (+95% and +92% respectively) and /a/ is less frequent (-54%).

All consonants are found in ideophones as well, with the most frequent ones (>10%) being the voiceless non-continuants /p, t, k/ and /r/. The voiceless non-continuants /p, t, k/ are the most frequent initial consonants (all together they make up half of initial consonants), /r/ is only found in medial position, and the velar consonants /k, g/ realized [k<sup>ɿ</sup>, ŋ] are the most frequent final consonants (they make up about 70% of the final consonants). Some of these distributional facts will be discussed again in the remainder of the paper. In comparison with the rest of the lexicon, voiced non-continuants in general are less frequent (-75% from the expected distribution of consonant classes regardless of parts-of-speech).

#### 2.4. Nasality in ideophones

Nasality in ideophones mostly follows the same distribution as in the rest of the lexicon. In the following ideophones, nasality spreads from a voiced non-continuant to all voiced segments on the left:

(2) /kibok/ [kĩmbok<sup>ɿ</sup>]\*  
‘swallowing, desiring’

(3) /bedaḡ/ [mēnāŋ]\*  
‘sticking the tongue out’

A few ideophones do not follow the expectations regarding nasality. The voiced non-continuant in (4) and the /r/ in (5) and (6) are expected to be fully nasalized, but are not.<sup>14</sup>

(4) /dub/ [ndũm]  
‘extremely loud noise (e.g. thunder)’

(5) /biribibig/ [mirimimĩŋ]\*  
‘liquid dripping softly’

(6) /tairikikid/ [tairikikin]\*  
‘disappearing’

---

<sup>14</sup> Other ideophones seem to diverge from the regular application of nasality, as visible in the absence of the tilde on vowels and /w/. In the absence of audio recordings for these items, this inconsistency could also be simply attributed to too broad a transcription.

Finally, it is interesting to note that out of the nine ideophones comprising a glottal non-obstruent /h/, six show nasal vowels. This association is sometimes called “rhinoglottophilia” (Matisoff 1975).

- (7) /hã/  
‘moving apart, opening’

To summarize this section, the phonology of ideophones almost always conforms to the usual phoneme inventory and restrictions in distribution. The distribution of particular phonemes in certain positions cannot serve as a defining criterion to identify ideophones, but nevertheless hints at the likeliness of an item being an ideophone. This skewed distribution is congruent with the foregrounding strategy for marking Teko ideophones in Akita (2021)’s terminology. For instance, the words in (8) and (9) conform to the prototypical phonology of ideophones.

- (8) /kog / [kõŋ]  
‘snoring’

- (9) /turuk/ [turuk̚]  
‘stepping down’

### 3. Root structure

#### 3.1. *Introduction to the Teko root and word structure*

Attested syllabic structures are quite simple in Teko: CV (by far the most frequent), V, CVC and VC. The maximal word pattern is made of a sequence of open syllables, with a closed syllable allowed in morpheme-final positions only. Each vowel of a vowel sequence is the nucleus of a separate syllable.

The quantitative analyses in this section are based on the information about root structure coded in the *Ideophone Type database* for ideophones, and in the *Noun, Verb and Ideophone database* for nouns and verbs. For each feature observed, a 3-sample chi-square test for equality of proportions was used to compare the observed distribution with a uniform distribution. Then post-hoc pairwise comparisons were carried out with p-values adjusted for multiple testing based on Holm’s method.

### 3.2. Root structure

Teko ideophones comply with the general structure of Teko roots. They make use of all allowed syllabic structures and only those. As monomorphemic words, they also follow the general word pattern. Only two items within the *Ideophone type database* contain a consonant sequence, which elsewhere is normally the result of morphological make-up.<sup>15</sup> One of these exceptions is potentially the repetition of a shorter ideophone (for which I have no evidence) (10), while the other is debatably not an ideophone. The form in (11) is used to render a song. Aplonova et al. (2022) have shown that in West African narratives, forms used to render songs, music or foreign languages (that they call ‘pseudo-words’ following Idiatov 2005) differ structurally from ideophones.

(10) *saŋsaŋ*\*  
‘chewing’

(11) *zinzawa*  
‘singing’

We have just seen that ideophones show root structures that are consistent with the Teko linguistic system. Nevertheless, their internal structure shows some of the allowed features much more frequently than roots of other word classes. This makes particular forms more likely to be identified as ideophones and the whole set to be distinguishable as a class of roots. Table 4 compares the root structure of ideophones, nouns and verbs in Teko.

	monosyllables	initial V	final C	(CV) <sup>n</sup> .CVC
ideophones	44%	3%	68%	64%
nouns	6%	25%	34%	21%
verbs	16%	11%	58%	46%

**Table 4:** Root structure of Teko ideophones, nouns and verbs.

<sup>15</sup> Rose (2011) was mentioning the ideophonic form *mankurug* as being unexpected due to the root-internal consonant sequence. Since then, it has been analyzed as a sequence of ideophones, based on the independent attestation of the ideophone *kurug* with a close meaning.

Regarding monosyllabicity, the observed distribution is significantly different from a uniform distribution ( $\chi^2(2) = 140.56, p < .001$ ). Comparisons show that ideophones are statistically much more often monosyllabic than nouns ( $p < .001$ ) or verbs ( $p < .001$ ).

Regarding the presence of an initial vowel, the observed distribution is also significantly different from a uniform distribution ( $\chi^2(2) = 54.213, p < .001$ ). Ideophones start with a vowel statistically less often than nouns ( $p < .001$ ) or verbs ( $p = .004$ ).

As for the presence of a final consonant, the observed distribution is again significantly different from a uniform distribution ( $\chi^2(2) = 82.926, p < .001$ ). Ideophones end with a consonant statistically more often than nouns ( $p < .001$ ) or verbs ( $p = .031$ ).

To summarize the general findings on ideophone root structure up to now, ideophones are monosyllabic statistically much more often than nouns or verbs, start with a vowel less often than nouns and verbs, and end in a consonant more often than nouns and verbs. In more specific terms, ideophones more often than nouns or verbs follow a structure made up of a final closed syllable possibly preceded by a series of open syllables, all syllables having a simple consonant onset:  $(CV)^n.CVC$ , as shown in the last column of Table 4. The observed distribution is significantly different from a uniform distribution ( $\chi^2(2) = 123.41, p < .001$ ). Comparisons show that ideophones follow this structure statistically more often than nouns ( $p < .001$ ) or verbs ( $p < .001$ ). The following ideophones illustrate this prototypical structure.

(12) *seg*  
 ‘sitting’

(13) *tapug*  
 ‘diving’

### 3.3. Monovocality and internal /r/

Other structural properties are strikingly more frequent in ideophones than in nouns or verbs, as shown in Table 5 on the subsets of multisyllabic items of each class.

	monovocality	internal /r/only
ideophones	48%	33%
nouns	16%	6%
verbs	26%	8%

**Table 5:** Additional structural properties of ideophones, nouns and verbs.

The first property is monovocality, defined by the exclusive presence of a single quality of vowels throughout the root (disregarding nasality). This property is found in half of the multisyllabic ideophones, such as those in (14) and (15).

(14) *ponoŋ*  
‘going out’

(15) *korokokor*  
‘tearing’

The observed distribution of monovocality across parts of speech is significantly different from a uniform distribution ( $\chi^2(2)=52.428$ ,  $p<.001$ ). Ideophones are monovocalic statistically more often than nouns ( $p<.001$ ) or verbs ( $p<.001$ ).

The second property is the exclusive presence of /r/ as a consonant root-internally (occurring a single or multiple times), found in a third of the multisyllabic ideophones. This is exemplified in (16) and (17).

(16) *kirog*  
‘swallowing’

(17) *purig*  
‘extricating’

The observed distribution of the exclusive presence of /r/ as a root-internal consonant is significantly different from a uniform distribution ( $\chi^2(2)=52.428$ ,  $p<.001$ ). Comparisons show that ideophones display only /r/ as a consonant root-internally more often than nouns ( $p<.001$ ) or verbs ( $p<.001$ ).

These two properties of multisyllabic ideophones are found combined more often in ideophones than in nouns and verbs (Chi2 test,  $p<.05$ ). They are dealt with together in this section because they associate in ideophones more than expected from their distribution ( $\chi^2(1, N=99)=14.74$ ,  $p<.001$ ). Examples are (18) and (19). The literature on vowel harmony does not mention a special role of flaps in vowel harmony (van der Hulst & van de Weijer, Jeroen 2011).

(18) *kurug*  
‘handling with hands or paws (grasping, digging)’

(19) *pururu*  
‘group falling’

A hypothesis to make sense of this strong association would be that multisyllabicity, monovocality and the exclusivity of internal /r/ together result from a particular morphological process, a type of reduplication adding more syllables to a root but substituting a default /r/ consonant in the copy in lieu of the original onset consonant. There is no indication that r-sounds or liquids are regularly used as fixed consonants in reduplication, but at least this pattern is attested in Malak Malak (Birk 2015: 95-96) and Somali (Dhoorre & Tosco 1998).

Monovocality in ideophones has been interpreted as expressing the regularity of the event (Dingemans 2011). A further hypothesis would be that the use of a default consonant also participates in expressing this regularity. In the particular case of (20), while the plain form *kʷəg* may be used for a simple jump of a frog or a monkey, the more complex form *kʷərəg* refers to a sub-event (a step) in climbing a tree or in walking up stairs, for example. These steps are more complex events than simple jumps, and are furthermore often part of a sequence (itself rendered by the repetition of the ideophone). If this hypothesis was confirmed, this would mean that many items that have been up to now considered the basic form of ideophones would in fact be derived from a simpler form, unattested in the (limited) sources.

(20) *kʷəg, kʷərəg*  
'jumping up, climbing'

To summarize this section, the structure of ideophones is not aberrant within the Teko linguistic system but regularly shows some features that are particularly representative of this class of words. This is another contribution to the foregrounding strategy for marking Teko ideophones, following Akita (2021)'s terminology.

#### **4. Word-level prosody**

Ideophones are often presented as being set apart prosodically by pauses, pitch or non-modal phonation (see for example Dingemans & Akita 2017; Mihas 2012). They can also show a different stress pattern (Reiter 2011: 297; Mihas 2012). They are also said to be easily manipulated prosodically for expressivity (Samarin 2001: 333). For example, Reiter (2011: 297-308) describes how variation in pitch, intensity and rhythm in Awetí ideophones can be used expressively in speech to modify the basic meaning of the ideophone.

The description of word-level prosody in ideophones presented in this section is based on a prosodic analysis of the 183 vowels from the items of the *Ideophone token database* (see *Vowel results*). Excerpts from audio recordings of discourse data were manually annotated in Praat (Boersma & Weenink 2023) for vowels and pauses before and after the ideophone (if need be).<sup>16</sup> Then the duration of the pauses and the pitch, mean intensity and duration of the vowels were extracted from them. Pitch was extracted at the beginning, end and at the first, second and third quartiles of the vowel length. The results, made available in the *Vowel results* sheet, were systematically compared to the analysis by Gordon & Rose (2006) of regular discourse data, which included both content and function words.<sup>17</sup>

#### 4.1. Introduction to Teko prosody

Gordon & Rose (2006) investigate stress in Teko. Other aspects of Teko word-, phrase- or utterance-level prosody have not been investigated. The domain of stress is the prosodic phrase. Primary stress usually falls on the penultimate syllable of the phrase or on the final syllable if heavy. Secondary stress alternates on every second syllable counting backward from the primary stress. Optional stress can also be found on the initial syllable of the phrase. The major acoustic correlates of stress are duration and intensity on words in isolation, and additionally pitch in discourse data (Gordon & Rose 2006).

#### 4.2. Stress in ideophones

A preliminary study on stress placement in ideophones indicates that it does not follow the general rules of stress placement in Teko, and is irregular across the word class.<sup>18</sup>

On disyllabic ideophones, stress either falls on the initial syllable (even when the final one is heavy as in [ˈdirikʔ] ‘watching’) or on the second one as in [peˈtekʔ]

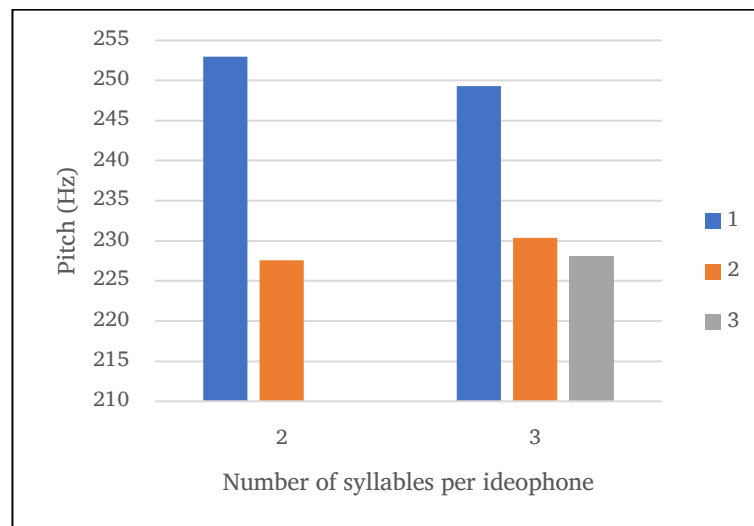
---

<sup>16</sup> The smallest pause is 27 milliseconds. Importantly, absence of pause segmentation in Praat was not coded as a zero-millisecond pause for the computation of the median duration of pauses.

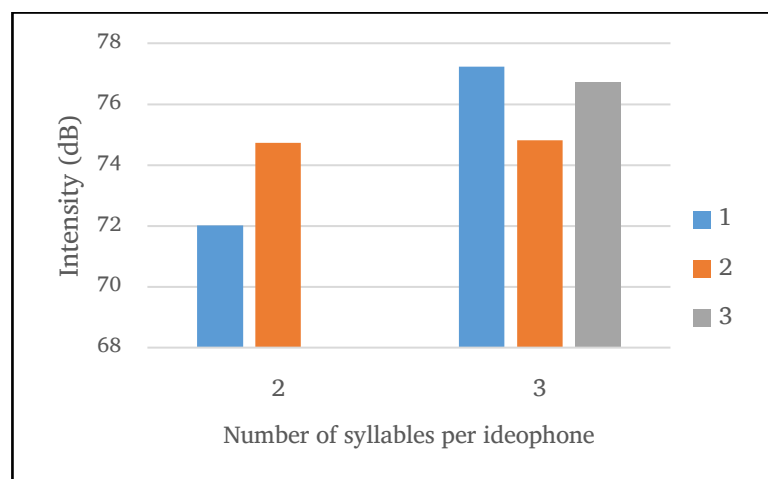
<sup>17</sup> Both studies exclude vowels in hiatus and vowels in word absolute final position, as these latter “were often characterized by a gradual shift into non-modal phonation (breathiness or creakiness) that made it difficult to determine their endpoints” (Gordon & Rose 2006). Additionally, nineteen vowels were excluded from our pitch analysis, because the items were glottalized.

<sup>18</sup> To check the transcription of stress placement, a sub-sample of 18 tokens of disyllabic and 6 tokens of trisyllabic ideophones have been submitted to a set of 5 transcribers, showing strong agreement for some items and lack of consensus on others.

‘pushing’, even when this is light as in [hi'ja] ‘walking’. Trisyllabic ideophones are stressed on their initial syllable, as illustrated in [ˈtəɾəɾəkʷ] ‘being noisy’. The following figures show the acoustic cues for stress in disyllabic and trisyllabic ideophones. Pitch and intensity are responsible for some prosodic saliency on the initial syllable, while duration highlights the penultimate syllable, as shown in Figure 1 to Figure 3.

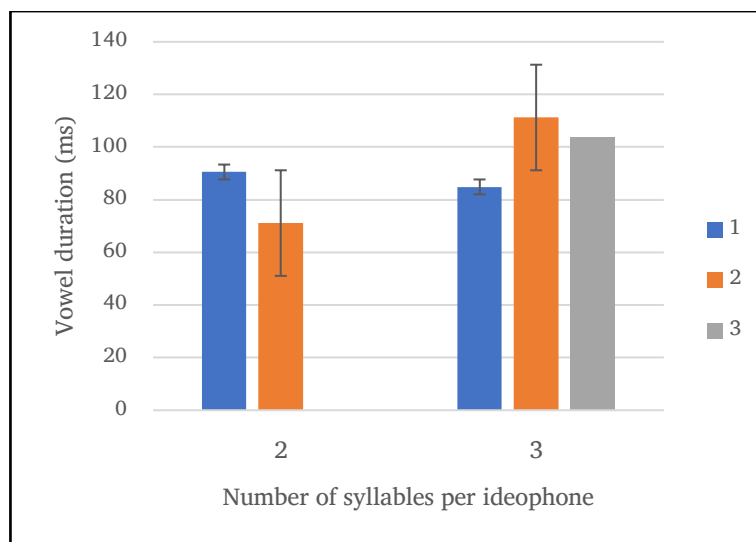


**Figure 1.** Pitch in disyllabic and trisyllabic ideophones.



**Figure 2.** Intensity in disyllabic and trisyllabic ideophones.





**Figure 3.** Vowel duration in disyllabic and trisyllabic ideophones.

It is for now unclear whether the unexpected stress placement described in this section as well as inconsistency within the word class could be explained by expressiveness overriding regular rules. We will see next how each prosodic cue compares in ideophones and other words, and can be recruited for expressive purposes.

### 4.3. Vowel duration in ideophones

The mean vowel duration of vowels in ideophones is in general comparable to the duration of vowels with primary stress in the discourse data, as shown in Table 6. Vowel lengthening is salient in monosyllables.

<i>ideophone token database</i>				Regular discourse data (Gordon & Rose 2006)		
all vowels	monosyllables	disyllables	trisyllables	primary	secondary	unstressed
96	183	83	100	99	85	84

**Table 6:** Average vowel duration (ms).

It is known from the literature on ideophones that the expressive prolongation of vowels is an iconic means of signifying an extension in space or time (Childs 1994; Reiter 2011). This expressive lengthening is for instance noticeable in the duration (955ms) of the final vowel of an occurrence of *pururu* ‘moving collectively’, probably expressing in that case the temporal and spatial extension of people falling from a

high tree due to height and number of people. Nevertheless, it is important to note that expressive lengthening is not limited to ideophones in Teko. Extended vowel duration is found with similar characteristics in parts-of-speech other than ideophones, such as the verb form *ohoŋ* ‘they are going’, in which the lengthening of the final vowel (857ms) expresses as well the extension of the motion event of two brothers climbing a high tree. In an example of *moŋ* ‘being dark’, the vowel is extremely long (574ms), and expresses the intensity of this sensory property. It can be compared to the salient lengthening of the attributive *epi* ‘it is expensive’ that often heard in the community as an expression of high intensity.

#### 4.4. Pitch in ideophones

In general, vowels in ideophones show a higher pitch than the surrounding text, as shown in Table 7 (note that the gender of speakers has not been normalized). As was mentioned above, pitch is higher on the first syllable in ideophones and then decreases throughout the word (Figure 1).

<i>Ideophone token database</i>	Regular discourse data (Gordon & Rose 2006)		
all vowels	primary	secondary	unstressed
242	165	168	159

**Table 7:** Average vowel pitch (Hz).

Some items show particularly remarkable variation in pitch. Salient variation in pitch is found in a similar fashion elsewhere in the language, for example in interjections.

#### 4.5. Intensity in ideophones

On average, vowels in ideophones show a higher intensity than the surrounding text, as shown in Table 8.

<i>Ideophone token database</i>	Regular discourse data (Gordon & Rose 2006)		
all vowels	primary	secondary	unstressed
75.9	67.8	66.2	65.5

**Table 8:** Average vowel intensity (dB).



verb, without their coda (Rose 2005).<sup>19</sup> Monosyllabic reduplication expresses event-internal repetition (including the plurality of participants, as in (22)), while disyllabic reduplication expresses either event-internal or event-external repetition as in (23) (Rose 2007, based on Cusic 1981).

(22) *amõ kito-kom õ-hẽ-hem.*  
 other frog-PL 3-RED-leave  
 ‘The other frogs leave.’ 13.045

(23) *õhẽ-õ-hem = ne o-ʔa.*  
 RED-3-leave = CONTR 3-fall  
 ‘He leaves again and falls.’ 21.237

## 5.2. Absence of morphology on ideophones

Teko ideophones do not normally combine with affixes or clitics. Nevertheless, this may not be a strict rule.<sup>20</sup> Morphology is at best a rarity on ideophones.

The absence of obligatory morphology and of any prefix in general makes ideophones highly distinguishable from other lexemes. However, similarly to syllabic and phonological structure, the absence of morphology is not sufficient to identify ideophones. But the presence of morphology (with the exception of valency-changing morphology, see 5.4) is sufficient to exclude the identification of a word as an ideophone. Indeed, when a root which shows many prototypical structural properties of an ideophone such as *karag* (a CVCVC structure, monovocality, internal /r/, a voiceless non-continuant in initial position and a voiced continuant in final position) takes a person prefix because it is a verb or a noun (as in *o-karag* ‘s/he fries’ for example), it diverges from the morphological patterns typical of ideophones.

<sup>19</sup> For monosyllabic verb roots, the domain of disyllabic reduplication also includes the person prefix, as in (22).

<sup>20</sup> The dictionary gives some examples of ideophones with what seems to be bound morphology, but their word boundary criteria are not always congruent with mine. Additionally, I have noted during an elicited session two examples of an ideophone with the interrogative clitic and two with the past clitic. In some published Teko tales (Renault-Lescure et al. 1987), some ideophones are followed by the form *iwər*, which is left untranslated and which I therefore suspect of being a discourse clitic. In the absence of recordings and specific work on this structure, I can only hypothesize that this *iwər* form could be a reduced form of *eʔi = iwər* ‘3S.say = CL’ involving the clitic = *iwər*, a well-identified clitic whose function is still unclear (see Section 6.2.3 on the verb *eʔi*). Note that the clitic would then not be attached to the ideophone but to the verbum dicendi (see Section 6.2.3).

### 5.3. Ideophone derivation

It is common to find that several ideophonic words can be related formally and semantically. I consider that this situation results from derivational processes affecting ideophone roots through duplication. Derived ideophonic forms are given in the column “Derived ideophonic form” of the *Ideophone type database*.

The two major processes of derivation are monosyllabic and disyllabic reduplication, disregarding the coda consonant and with the domain of reduplication starting from the left edge, just as in verbal reduplication (see Section 5.1).<sup>21</sup> Reduplication in ideophones may involve other formal modifications unattested in verbal reduplication: the substitution of the original consonant of the reduplicand with /r/ in the copy, as already discussed in Section 3.3, the addition of a vowel word-finally, or the “triplication” of the reduplicand (with a double copy). These different derivation processes are illustrated in Table 9.

ideophone root	meaning	derived stem	meaning	reduplication	modification
<i>pitiŋ</i> *	shuddering	<i>pipitiŋ</i> *	intense shaking	monosyllabic	
<i>kʷəg</i>	jumping	<i>kʷərag</i>	climbing	monosyllabic	C → /r/
<i>tig</i> <sup>*22</sup>	walking (stepping?)	<i>tititig</i>	group running	monosyllabic	triplication
<i>durug</i>	standing up	<i>dururug</i> *	standing up	monosyllabic	copy of rV non-initial syllable/ or triplication with C → /r/
<i>tir</i>	having buttock up	<i>titiri</i> *	having buttock up, back arched	monosyllabic	addition of final vowel
<i>kitig</i> *	shaking (head, body)	<i>kitikitig</i> *	nonstop shaking (head, body)	disyllabic	

**Table 9.** Examples of ideophonic derivation.

<sup>21</sup> Because ideophones are never prefixed, the domain of reduplication is restricted to the initial syllables of the root (compare with footnote 19).

<sup>22</sup> The basic form is not found in my corpus, nor in the dictionary, but the repeated form *tigtig* is given in the dictionary with the ‘walking’ meaning (Cachine et al. 2020: 102).

As far as the semantics associated with these various derivational processes are concerned, they pertain to pluractionality, expressing either plurality of participants, event-internal repetition, increase in duration which turns a punctual event into an activity, increase in intensity, or several of these (mostly aspectual) meanings. There are not enough data to specify whether specific meanings are associated with specific derivational patterns.

#### 5.4. Word-class changing derivation

The grammar of Teko (Rose 2011: 403) asserted that ideophones could not be the result nor the source of derivation. However, the Teko dictionary (Cachine et al. 2020) provides precious information on word-class changing derivation processes involving ideophones, by having non-ideophones derived from ideophones listed as sub-entries of ideophone entries.

First of all, deideophonic conversion to several major lexical classes has been observed. Potential cases are listed in the column “Word-class changing derivation” of the *Ideophone type database*,<sup>23</sup> and examples are provided in Table 10, sometimes showing some formal modifications typical of derived ideophonic forms as described in Section 5.3. Deideophonic roots follow the regular morphosyntactic behavior of their parts-of-speech.

ideophone	meaning	deideophonic root	meaning	part-of-speech
<i>wur</i> *	moving up	<i>wur</i> *	move up	intransitive verb
<i>susug</i>	suckling	<i>susu</i>	breast	noun
<i>worog</i> *	feeling a gap in the ground	<i>woroworog(i)</i> *	be soft (for the ground)	attributive

Table 10. Examples of deideophonic conversion.

Second, the dictionary regularly provides verb stems derived from ideophones through valency-changing morphology, such as the causative *bo-* [*bo* ~ *mō*] as in (24) or the middle *ze-* as in (25). The resulting deideophonic stems follow the regular morphosyntactic behavior of verbs.<sup>24</sup>

<sup>23</sup> A question mark within parentheses indicates less straightforward cases.

<sup>24</sup> Note that, as will be discussed in 8.3, ideophones lack intrinsic valency. There is therefore no use for valency-changing morphology on ideophones.

- (24) *puteḡ o-mo-kosoŋ.*  
 bottle 3-CAUS-IDEO.shaking  
 ‘S/he is shaking the bottle.’ (Cachine et al. 2020: 62)

- (25) *-ze-kũ-menəŋ*  
 MID-tongue-IDEO.sticking\_out\_tongue  
 ‘Stick the tongue out’ (Cachine et al. 2020: 72)

Actually, Teko valency-changing morphology has long been recognized for its potential to combine with various classes of roots, including verbs and nouns (Couchili et al. 2002; see also chapter X of Rose 2011). In all cases, the resulting predicates behave morphologically as verbs. Consequently, ideophones constitute a regular lexical class, subject to verbalization.

## 6. Syntax

### 6.1. Introduction to the syntax of Teko

Teko can best be described as a verb-final language. The predicate is the only obligatory constituent. The subject is normally placed before the predicate, and the object may precede or follow it. Adverbs and subordinate clauses are clause initial as in (26).

- (26) *pitaŋ o-kiḡe-r-ehe, takuru-ṽũwĩ*  
 child 3-be\_scared-RELN-because rock-DIM  
  
*o-tui-o-maṽe-r-ehe o-wur.*  
 3-be-CONT-REL-RELN-on 3-go\_up  
 ‘Since the child is scared, he goes up on a small rock.’ 13.028

Any kind of constituent may be focused by being dislocated to the clause initial position, where it carries second position clitics with various discourse functions (like focus = *te* and interrogative/exclamative = *sipo* in (27)).

- (27) *e-men-a = te = sipo iḡe a-ṽu [...]!*  
 1SG-husband-REF = FOC = INTER/EXCL PRO1SG 1SG-eat  
 ‘This is my husband I ate!’ 21.181

## 6.2. Syntactic integration of ideophones

I distinguish four types of syntactic integration, presented in the following subsections: holophrastic, collocational, light-verb argument, argument-taking. These four types are presented in Table 11 from the least to the most integrated with the following stretch of speech. In the first two levels, the ideophone is completely independent syntactically. It is optional and has no syntactic interaction with the following clause: it can be considered an extra-clausal element. In the other two levels, the ideophone is part of the clause syntax and cannot be deleted without making the clause ungrammatical.

	status	tokens/101
holophrastic	independent	23
collocational		59
light-verb argument	integrated	9
argument-taking		10

**Table 11.** Levels of syntactic integration of ideophones.

Ideophones are much more frequently found in their independent uses than integrated, as shown by the figures from the *Ideophone token database* given in Table 11. Note that the same individual ideophone can occur with different levels of syntactic integration (compare for example (28) and (34)).

### 6.2.1. Holophrastic

The holophrastic use of ideophones covers cases where the ideophone has no syntactic interaction with other elements, and is the sole element to inform on a particular event. It is syntactically optional, in that the rest of the surrounding speech would still be grammatical without it. Nevertheless, it is not optional semantically: its content is highly informative. All in all, it could be considered to form a clause on its own.

In (28), *dirig* depicts a ‘watching’ situation, which is not expressed elsewhere in the utterance. In the context, the main character of the myth hides near a village to find out (through observation) who is the person cooking for him in secret. He then sees a woman coming out of a monkey skin. The ideophone here makes up a whole clause.





(32). They can be preceded by extra-clausal elements only, like discourse connectors, such as *ko* in (33).

(31) *surug*                      *surug*                      *aʔe-koti = ne = ʔe*                      *o-nenan-ɔŋ.*  
 IDEO.entering              IDEO.entering              DEM-at = CONTR = INTENS              3-enter-PL.S  
 ‘They enter at her place.’ 21.117

(32) [...]*pitij*                      *tapug*                      *o-por = eʔe*                      *ʔi-b*                      *o-ʔar-ɔŋ.*  
 IDEO.shuddering              IDEO.diving              3-dive = INTENS              water-in                      3-fall-PL.S  
 ‘Splash, they dive into the water’ 04.042

(33) *ko*                      *kir*                      *o-wir*                      *o-ʔur.*  
 then                      IDEO.going\_off                      3-come.off                      3-come  
 ‘Then, splash, she falls’ 05.032

Semantically, a collocational ideophone contributes to the semantics of the stretch of text less crucially than a holophrastic ideophone. Indeed, it expresses the same event than the verb it collocates with, but with some additional specification in terms of manner, type of participant, or aspect (See Section 8). Most ideophones regularly collocate with the same verbs (column “Verb in collocation” in *Ideophone type database*). This is for example the case of *dirig*, found nine times in a collocational use in the texts, each time with the same verb *maʔẽ* ‘watch’. Example (34) illustrates this frequent collocation.

(34) *dirig*                      *o-maʔẽ-katu*                      *ikeʔi*  
 IDEO.watching              3-watch-good                      so  
 ‘So she’s watching well’ 21.180

### 6.2.3. Light-verb argument

In the light-verb construction, the ideophone is introduced by a light verb. These light verbs are generally *ʔe* ‘say, make (a noise)’ or *baʔe* ‘do, make’, more rarely *tui* ‘be, become’.<sup>25</sup> It has been long recognized that ideophones are often introduced with the verb used in quotative constructions (Güldemann 2008: 280-283). Light verb *ʔe* is illustrated in (35), and *baʔe* in (36).

<sup>25</sup> The verb *ʔe* has an irregular form *eʔi* with a third person subject, and *ere* in the imperative with a singular subject.

(35) *nan = āhã*                      *miŋ*                      *eʔi-o*  
 like\_this = only                      IDEO.closing\_eyes                      3.say-CONT  
 ‘She (a dead person) closes the eyes then.’ 21.044

(36) *nɪŋ*                      *o-baʔe*                      *e-koti.*  
 IDEO.winking                      3-do                      1SG-toward  
 ‘He blinked at me.’ elicited

The ideophones immediately precede the light verbs, in a usual position for an object. The verb *ʔe* normally introduces strictly pre-verbal direct speech, as in (37), more rarely a pre-verbal pronominal object as in (38). The verb *baʔe* normally introduces an object noun phrase, which is often but not always pre-verbal (39) and sometimes marked by a translative case marker (40).<sup>26</sup> Ideophones introduced by *baʔe* are always preverbal and never case-marked.

(37) *am*                      *wane = so*                      *eʔi.*  
 here                      fine = INTER                      3.say  
 ‘“Is it fine here?” he asks.’ 01.021

(38) [...]*mati = sipo*                      *nan*                      *eʔi?*  
 where = INTER/EXCL                      thus                      3.say  
 ‘Where does this sound come from?’ 21.161

(39) *kija*                      *o-baʔe.*  
 hammock                      3-do  
 ‘She makes a hammock.’ 06.024

(40) *o-baʔe*                      *pari-am-oŋ.*  
 3-do                      boucan-REF-PL.S  
 ‘They made a boucan’. 11.025

An ideophone introduced by a light verb is required for the clause to be grammatical. It is internal to the clause, as shown by the presence of a second-position clitic = *āhã* on the constituent *nan* preceding the ideophone in (35).

<sup>26</sup> The translative case indicates a change of state for the referent of the noun, from non-realized to realized (Rose 2011: 235-240).

In the sources, some ideophones are attested with only one of the light verbs, some with *?e* only (such as *sa* ‘small rain’), others with *ba?e* (such as *kar* ‘cut’) and still others with *tui* (*mi?ũmi?ũ\** ‘feeling internal contraction’). However, the same ideophone can combine with several light verbs, as illustrated by the comparison of (41) and (42).

(41) <i>moŋ</i>	<i>e?i</i>	(42) <i>moŋ</i>	<i>o-ba?e</i>
IDEO.darkness	3.say	IDEO.darkness	3-make
‘It is dark.’		‘S/he turns the light off’	

In general terms, *ba?e* is used when the subject participant exerts some control on the state of affairs expressed by the ideophone, while *?e* does not imply control.<sup>27</sup> This semantic contrast can be observed in examples (41) and (42), as well as (35) and (36). As a consequence, *ba?e* is almost exclusively used with animate participants. In contrast, *?e* is used with participants of the following types: impersonals as in (41), inanimates (43), animals (44), dead humans as in (35), or living humans with ideophones expressing inner feelings (45), physical reflexes (46),<sup>28</sup> and body actions that involve some control over one’s body (47). This distribution reminds us of the connection between ideophones and the middle functions put forward by van Gijn (2010) on the basis of Yurakaré data. Noticeably, imperative clauses often unexpectedly make use of the *?e* light verb (see (58)).

(43) <i>tif</i>	<i>tif</i>	<i>e?i</i>	<i>i-ka-wər</i>	<i>pari-?ar-o.</i>
IDEO.fat_dripping	IDEO.fat_dripping	3.say	3-fat-?	boucan-on-CONT
‘Their fat does tish tish (when it drips) on the boucan (wooden grill).’ 37.090				

(44) <i>zawar = enam</i>	<i>ka-r-ehe</i>	<i>wog</i>	<i>wog</i>	<i>wog</i>	<i>e?i.</i>
dog = TOP_SW	wasp-RELN-at	IDEO.barking	IDEO.barking	IDEO.barking	3.say
‘The dog barks to the wasps: "wow, wow."’ 16.018					

<sup>27</sup> In one example, the ideophone actually depicts the state of affairs of the human object participant of *ba?e*, the subject of which is inanimate. Here *ba?e* seems to have been selected for its causative sub-component, even though the subject does not exert any control over the object.

<i>Wiu</i>	<i>mōāhā</i>	<i>kuku</i>	<i>e-ba?e.</i>
IDEO.being_dizzy	like_this	manioc_beer	2-make
‘The manioc beer made me dizzy.’ (Cachine et al. 2020: 119-120)			

<sup>28</sup> My corpus shows several exceptions to this with the ideophones *pur* and *tifag* for ‘farting’ being introduced with *ba?e*.

- (45) **Ari** *eʔi = ne = te.*  
 IDEO.feeling\_pity 3.say = CONTR = FOC  
 ‘He had pity.’ (Cachine et al. 2020: 6)
- (46) **Atug** *a-ʔe-tanẽ-ãbit* *iɕe.*  
 IDEO.vomiting 1SG-say-DESID-ASSERT 1sg  
 ‘I feel nauseous.’ (Cachine et al. 2020: 10)
- (47) **Menaj** *eʔi.*  
 ideo.sticking\_out\_tongue 3.say  
 ‘He sticks his tongue out.’ (Cachine et al. 2020: 72)

When ideophones are introduced by the light verb *tui* ‘be, become’, it seems this adds an inchoative aspect to the event depicted by the ideophone, as in (48) and (49). The notion of control is irrelevant.

- (48) **Wiu** *mõãhã* *e-akaŋ* *o-tui.*  
 IDEO.being.dizzy like.that 2-head 3-be  
 ‘My head spinned all the sudden.’ (Cachine et al. 2020: 119)
- (49) **tir** *tir*  
 IDEO.with\_arched\_back,\_buttock\_up IDEO.with\_arched\_back,\_buttock\_up  
*o-u-koti* *o-tui.*  
 3-father-towards 3-do  
 ‘They showed their butts to their father.’ 32.067

#### 6.2.4. Argument-taking

In its argument-taking use, the ideophone plays the structuring role of a predicate. There is no verb co-expressing the event. The ideophone governs preceding or following patients or oblique arguments, as in examples (50) to (52). In all the examples from the *Ideophone token database*, the subject is inferred from the context or the construction (see for example the prohibitive construction in (55)).

- (50) **peteg** *o-kuŋa* *t-o-ʔar* *o-ho.*  
 IDEO.pushing 3-brother PURP-3-fall 3-go  
 ‘He pushes his brother to make him fall.’ 21.024

(51) *ponoŋ* *i-koti* *kiʔi*  
 IDEO.arriving,moving\_out 3-toward so  
 ‘So he goes out toward him.’ 21.225

(52) *siriɖ* *waita,* *siriɖ* *sapa.*  
 IDEO.grasping basket IDEO.grasping machete  
 ‘Hop, a basket, hop, a machete.’ 22.200

Remarkably, in a construction where the ideophone takes arguments, second-position clitics still attach to the element following the ideophone. The absence of morphological potential of ideophones is maintained even in cases of tight syntactic integration.

(53) *siriɖ* *i-zebiʔa = ne = ʔe* *siriɖ* *i-zebiʔa.*  
 IDEO\_grasping 3-music = CONTR = INTENS IDEO\_grasping 3-music  
 ‘He takes his music.’ 21.149

### 6.3. Syntactic integration and prosodic integration of ideophones

To study whether the types of syntactic integration sketched above are supported by prosodic integration, pause duration was observed for the items of the *Ideophone token database*. The results are given in the last two columns of the *Ideophone token database*.

First, the pause before ideophones was examined, excluding ideophones that are preceded by another ideophone. Forty-five out of the remaining 56 ideophones (80%) are preceded by a pause, and the median duration of the pause is 753ms. This long duration is not surprising given the fact that most ideophones are found clause-initially. In fact, this number is not too far from the median duration of inter-clause pauses (993ms).<sup>29</sup>

Second, the pause after ideophones was examined, this time excluding ideophones which are followed by another ideophone. In Table 12, the results were split depending on the four levels of syntactic integration sketched above.

---

<sup>29</sup> This number has been estimated on the basis of a single but long text of about 20 minutes.

	status	presence of pause	presence of pause	median duration of following pause in ms
holophrastic	independent	4/6	67%	422
collocational		17/35	49%	163
light-verb argument	integrated	2/7	29%	143
argument-taking		1/8	12,5%	090

**Table 12.** Association of prosodic and syntactic integration of ideophones.

Table 12 shows a clear association between the prosodic and syntactic integration of ideophones: the more syntactically integrated ideophones are (i.e. from the top to the bottom rows), the less they are followed by pauses and the smaller the pauses.<sup>30</sup>

Ideophones are rather well-integrated prosodically in the light-verb construction, as in direct speech with the quotative verb (Rose & Vanhove 2007) and in the argument-taking construction (as in VPs), i.e. in the two constructions from which they cannot be deleted without making the clause ungrammatical. On the contrary, ideophones more often form a separate prosodic unit when they are optional and morphosyntactically independent: this supports their analysis as clause-external and does not support the idea that collocational ideophones could be syntactically “adverbial”. The present study on pauses and syntactic integration of ideophones nicely complements that on the inverse relation between expressiveness and syntactic integration of ideophones, based on pitch, phonation type, reduplication & lengthening (Dingemanse & Akita 2017): the former focuses on the framing strategy, and the latter on the foregrounding strategy for marking ideophones, in Akita (2021)’s terms.

To summarize this section, it is important to highlight that Teko ideophones show varying degrees of syntactic integration. They are not necessarily extra-clausal.

## 7. Discourse use

The pioneering literature on ideophones suggested that ideophones were expected to occur in limited discourse environments, due to their expressive nature. For example, Childs (1994: 194-195) stated that African ideophones were generally found in declarative sentences and genres associated with performance. More recent literature

<sup>30</sup> The presence and duration of pauses after ideophones in different syntactic integration patterns cannot be robustly assessed statistically given the small number of tokens in three of the four categories.

accounts for the use of ideophones in everyday speech (for example Dingemanse 2011; Mihas 2012) as well as in various special verbal arts (a list is given in Dingemanse 2012: 665). As for sentence types and negation, it still seems to be commonly assumed that ideophones do not combine easily with interrogation and negation (Kita 1997: 389-391; Kilian-Hatz 2001: 158; Reiter 2011: 355; Dingemanse 2012: 667).

The use of Teko ideophones is certainly not exclusively restricted to some sentence types (7.1) or genres (7.2), but it shows preferences in its distribution. This section will also describe how ideophones often come in series (7.3).

### 7.1. Sentence types

Teko ideophones are not restricted to declarative sentences. The following examples show ideophones in interrogative (54) and imperative (55) sentences.

(54) *seg*                      *mõ*                      *s = o-apig ?*  
 IDEO.sitting    like\_that    INTER = 3-sit  
 ‘Did s/he sit like that?’ elicited

(55) *mame*    *tərərəg*                      *eʔi*                      *zawar-a-pe.*  
 NEG.IMP    IDEO\_making\_noise    3.say                      dog-REF-to  
 ‘“Don’t make noise”, he tells the dog.’ 16.042

Ideophones are therefore not just used to depict reality. They can indeed be used for non-realized events, as additionally evidenced by their use in negative clauses such as (56). Examples (56) and (57) also show that ideophones can be used to refer to a state of affairs involving the speech participants.

(56) *kor*                      *tamadua*                      *miŋ*    *d-eʔi-ɕi.*  
 then                      giant\_ant eater                      IDEO    NEG-3.say-NEG  
 ‘Then the giant anteater didn’t close its eyes.’ 29.009

(57) *miŋ*                      *si-ʔe-tar-te-eʔe*                      *si-posi-ŋ.*  
 IDEO.closing\_eyes    1INCL-do-FUT-CL    1INCL-shit-CONT  
 ‘Let’s do our business with our eyes closed.’ 29.005





Narratives were further categorized into tales (myths and animal fables), historical narratives, personal narratives and rendering of the picture book *Frog, where are you?* (Mayer 1969). The proportion of ideophones in tales was then compared to that in other types of narrative all together in Table 14.<sup>32</sup> A 2-sample chi-square test for equality of proportions indicates that the proportion of ideophones in the traditional tales of the corpus is significantly greater than for those in other types of narrative ( $\chi^2(1) = 459.81, p < .001$ ). It seems ideophones contribute to the performance which determines the quality of the telling of a traditional tale.

	ideophones	words	percentage
tale	364	4,709	8%
other narrative	54	3,327	1.6%

Table 14. Ideophones in different types of narrative texts.

### 7.3. Series of ideophones

Ideophones in natural speech often come in series. Out of the 101 items of the *Ideophone token database*, 45 had been uttered within series of two or more ideophones. Within those, 15 were followed by a different ideophone, and 30 by the same ideophone.

#### 7.3.1. Series of different ideophones

Series involving different ideophones from the *Ideophone token database* involve two to three ideophones, rendering a sequence of events in chronological order. For instance, the series of three ideophones in (59) depicts three successive events involving a turtle stuck in a tree: its shaking to get free, its detachment from the tree, and its fall.

- (59) *kor kuʔe-kuʔe kir tou*  
 then RED-IDEO.wriggling IDEO.going\_off IDEO.falling  
*o-wir o-ʔar zawapinim-a-ʔar.*  
 3-go\_off 3-fall leopard-REF-on  
 ‘Then it moves and falls down on the leopard.’ 05.025

<sup>32</sup> The ‘other narrative’ category combines 46 ideophones from 2,209 words of historical tales, 8 ideophones from the 645 words of Frog stories, and the absence of ideophones within the 527 words of personal narratives.

7.3.2. Series of identical ideophones

Series involving identical ideophones from the *Ideophone token database* involve up to eight repetitions, and possibly include some items in a derived form (see Section 5.3). Semantic effects of ideophone repetition pertain to pluractionality: either plurality of participants as in (60), or plurality of events as in (61).

(60) *siriɔ*                      *siriɔ*                      *pakuʔa* [...] *o-ho*                      *i-koti-ŋ* [...].  
 IDEO.grasping IDEO.graping banana                      3-go                      3-at-PL.S  
 ‘They take the bananas and go to his place’ 04.007

(61) *pau*                                      *pau*                                      *mokoŋ*                      *apapu*                      *iwər*  
 IDEO\_rifle\_shooting IDEO\_rifle\_shooting two                      rifle\_shooting ?  
  
*o-kaŋum*                      *o-ho*.  
 3-disappear                      3-go  
 ‘He shot two gunshots and disappeared far away.’ 23.065

Event packaging seems tighter in series of identical ideophones than in series of different ideophones (Table 15). A Mann-Whitney U test (aka Wilcoxon rank-sum test) shows that the pauses are significantly longer when ideophones in series are different (Mdn = 0.17) than when there are identical (Mdn = 0.09),  $U = 231, p < .001$ .

	number of pauses /tokens	percentage of pauses	median duration of following pause (ms)
different	10/15	67%	171
identical	29/30	96%	091

Table 15. Ideophones in series.

8. Semantics

The understanding of the semantics of Teko ideophones is based on the comparison of their uses in the text corpus, discussion during work sessions with consultants, as well as the definition and examples given in the published dictionary (Cachine et al. 2020). As this understanding is highly dependent on the limited amount of data

available, it should be taken as preliminary. It is indeed likely that the actual use of individual ideophones is either more specific or more general than supposed.

The *Ideophone type database* gives a provisional gloss (column “Provisional gloss”), and categorizes each ideophone along sensory modalities, Levin’s verb classes and various components of motion semantics (the following four columns). These three attempts to categorize the meaning of ideophones follow different approaches. The first one discussed in Section 8.1 follows a categorization in sensory modalities supposed to be highly relevant for ideophones (Dingemanse 2012: 663). The second approach, followed in 8.2, is a categorization of state of affairs expressed by verbs in English (Levin 1993). This categorization is one of the few comprehensive lists of events available in the literature, and I have taken it as a categorization that would not be biased by our prior knowledge of ideophones, even though I am fully aware that it is itself biased by the morphosyntax and semantics of English verbs. The third categorization, discussed in 8.3, is a typologically-oriented coding system specific to ideophones expressing motion events (Ibarretxe-Antuñano 2019), based on a long history of semantic work on the domain of motion since Talmy (2000)’s pioneering work. After the semantics of Teko ideophones is investigated through these different approaches, Section 8.4. discusses potential phonosemantic characteristics of Teko ideophones.

Before this, a few general remarks are in order. From working sessions with my main consultant, it is clear that ideophones are part of the standard lexicon of the language. They are easily identified as belonging to a particular class of words and their meaning is conventionalized. As such, the meaning of an ideophone can be discussed even when cited in isolation. My consultant is able to offer varied utterances including the investigated form, and referring to diverse situations covered by this form. As a final general note, the depictive function of ideophones is sometimes overtly signaled in discourse by the use of the manner adverb *moŋi* or *moãhã* ‘this way, like that’ as illustrated in (62).

- (62) *fu*                      *moãhã*                      *node-peçũ*                      *a?e-a-te*                      *node-ba?e*.  
 IDEO.blow                      like\_that                      1INCL-blow                      this-REF-FOC                      1INCL-do  
 ‘He blew us, this is what made us.’ 02.021

### 8.1. *The sensory modalities hierarchy*

On the basis of previous publications, some generalizations have been made as to which semantic areas are usually covered by ideophones. These have been summarized in an implicational hierarchy, presented in Figure 4.<sup>33</sup>

SOUND < MOVEMENT < VISUAL PATTERNS < OTHER SENSORY PERCEPTIONS < INNER  
FEELINGS AND COGNITIVE STATES

**Figure 4.** Sensory modalities implicational hierarchy for ideophones (following Dingemanse (2012: 663)).

The implicational hierarchy should be read as follows: “if a language has ideophones at all it will have at least ideophones for sound (i.e. onomatopoeia). If a language has ideophones for movement it will also have ideophones for sounds. If a language has ideophones for visual patterns (e.g. spatial configuration or surface appearance), it will also have ideophones for movements and sounds, et cetera. Conversely, a language that does not have ideophones for sounds or movements will not have ideophones for cognitive states” (Dingemanse 2012: 663).

I have attempted to categorize Teko ideophones into the different levels of the hierarchy (see column “sensory modalities” of *Ideophone type database*) in order to assess the validity of this implicational hierarchy for Teko data. Two main difficulties were encountered. First, the categories in the hierarchy are not exclusive, as many ideophones can be said to depict both movement and sound. These were coded as depicting movements, while those that have been coded as depicting sounds clearly do not entail movement. Second, the categories are not exhaustive. Many ideophones indeed do not fall into any of the proposed categories, without having an unreasonably wide understanding of movement, or arbitrarily considering sound as their primary component. Consequently, an additional category labelled “action other than movement” was created. Most ideophones in this category depict actions to which a sound can often be associated, but not necessarily. Examples for all semantic

---

<sup>33</sup> McLean (2021) offers a revised version of this hierarchy in the light of Japonic data, in which the steps VISUAL PATTERNS < OTHER SENSORY PERCEPTIONS of the hierarchy are replaced by FORM < TEXTURE < OTHER SENSORY PERCEPTIONS. This proposal is not relevant for the present Teko data, in which no ideophone for “form” and “texture” has been documented.

areas are given in Table 16, as well as the number of ideophones from the list categorized in these areas.

SEMANTIC AREA	sound	movement	visual patterns	other sensory perceptions	inner feelings and cognitive states	action	others
COUNT	26	70	3	5	5	58	10
EXAMPLE	<i>wɔŋ</i>	<i>peteg</i>	<i>wo</i>	<i>tiriri</i>	<i>ari</i>	<i>pig</i>	<i>koɕ</i>
TRANSLATION	'whistling'	'pushing'	'light'	'slightly spicy'	'feeling pity'	'(un)covering eyes with hand'	'in a short time'

Table 16. Ideophones and semantic areas.

Table 16 shows that all semantic areas traditionally covered by ideophones are relevant to the analysis of Teko ideophones. Yet sound and especially movements are clearly dominant areas, while ideophones related to other senses are much less common. While this distribution does not contradict the implicational hierarchy, the fact that the additional ‘action’ category groups almost a third of the ideophones in the list points to the fact that the categories within the hierarchy are not cross-linguistically consistent.

### 8.2. Levin’s verb classes

Following the results from the previous section, it seemed important to focus on the action and movement categories and account for all types of events. The labels of the 49 classes of English verbs identified by Levin (1993) for English on semantic and morphosyntactic grounds were useful in offering a first categorization of Teko ideophones in terms of event type (see column “Levin’s categorization” of *Ideophone type database*), since most of them involve some dynamicity or change of state, just like verbs in English.

The list of 177 Teko ideophones can be distributed into 26 of Levin’s 49 classes, but only three classes are heavily populated: 20 ideophones for “emission”, 24 “involving the body”, and 46 “motion”. Other classes with at least five elements are

the following: “assuming a position”, “change of state”, “contact by impact”, “cutting”, “ingesting”, “perception”, and “removing”.

The three major classes of ideophones according to this categorization can be examined in detail and compared to the results of the previous section (a detailed subcategorization for these major classes is offered in column “Levin's subcategorization” of the *Ideophone type database*. Ideophones of emission mostly include emission of sound, but also light, and substance.<sup>34</sup> With the addition of five ideophones for “sounds made by animals”, they more or less correspond to sound, and visual patterns in the sensory modality classification. Ideophones “involving the body” include bodily processes, non-verbal expression, gestures/signs involving body-parts, body-internal states of existence and verbs of bodily state and damage to the body. This class does not correspond neatly to the sensory modality classification. Finally, the difference between the 46 ideophones of motion in this section and the 70 movement ideophones identified in the preceding section stems from the fact that the latter comprise not only motion semantics, but also caused motion and change of posture, which in Levin's finer categorization pertain to other classes.

### 8.3. *The motion semantic grid*

The motion semantic grid for ideophones developed by Ibarretxe (2019) proposes a number of variables to break down the motion semantics into components associated with the figure, the ground, the path, the manner, the cause, the event extension (phase or aspect), and whether the motion is movement (translocation) or stationary. All 70 ideophones coded as expressing movement in Section 8.1 have been coded for these components (the last 26 columns of the *Ideophone type database*). Two types of information can be extracted from this coding: which semantic components of motion are lexicalized in some ideophones, and which specific semes are found for each relevant component.

Regarding the figure of motion, only two motion ideophones clearly involve the number of the figure (*pororo* ‘group leaving and spreading out’, *pururu* ‘falling collectively’) while others are neutral in that respect. As an example, the ideophone *tapug* ‘diving’ depicts the diving of one or several bodies (compare (63) with (64)). Repetition of the ideophone is often used to express the plurality of participants (65).

---

<sup>34</sup> None of the Teko ideophones have been considered to express emission of smell, though the three ideophones for ‘farting’ could be analyzed as such. They have been coded as ‘bodily processes’ in the database.

Ideophones are generally neutral in terms of number of figure even when the verbs that the ideophone associates with lexicalize the number of participants, as the verbs *ike* ‘enter’ and *nenan* ‘enter as a group’ that are found in collocational constructions with the ideophone *surug* ‘entering’ (compare (66) with (67)).

- (63) *tapug*                      *iwər*                      *tihākã-pope*                      *o-mobor*.  
 IDEO.diving                      ?                      cove-in                      3-throw  
 ‘He threw (the creeper) in the cove.’ 24.013
- (64) *tapug*                      *ze-kapirer = ne*                      *ʔi-b*                      *o-po-por*                      *o-ho-oŋ*.  
 IDEO.diving                      RECP-behind = CONTR                      water-in                      3-RED-jump                      3-go-PL.S  
 ‘Splash, they dive in the water one after the other and go away.’ 04.035
- (65) *tapug*                      *i-pope = ne*                      *oʔa-o-ʔar-oŋ*                      *tapug*                      *tapug*.  
 IDEO.diving                      3-in = CONTR                      RED-3-fall-PL.S                      IDEO.diving                      IDEO.diving  
 ‘Splash, they go (lit. fall) in it (the stomach of the tapir), splash, splash.’ 04.024
- (66) *surug*                      *surug*                      *aʔe-koti = ne = ʔe*                      *o-nenan-oŋ*.  
 IDEO.entering                      IDEO.entering                      DEM-at = CONTR = INTENS                      3-enter-PL.S  
 ‘They go back home (to the Tebesig).’ 21.117
- (67) *surug*                      *moŋi*                      *bato*                      *o-ike*                      *o-ho*                      *ʔi-b*.  
 IDEO.entering                      this\_way                      boat                      3-enter                      3-go                      water-in  
 ‘The ship enters under the sea.’ 37.015

As far as the animacy of the figure is concerned, it is often neutralized in motion ideophones (and other ideophones as well). For instance, the diving process expressed by the ideophone *tapug* is realized by an inanimate participant, a creeper, in (63) and by animate participants, girls, in (64) . As for the entering process expressed by the ideophone *surug*, it can be realized by animate participants, like the boys in (66) or by an inanimate participant, such as a ship in (67). As noted in Rose (2011), the figure of *tou* ‘moving straight in the air, vertical or horizontal, until contact’ (often used for falling) in the sources is in turns a turtle, a monkey, people, and seeds, and the figure of *siriŋ* ‘grasping’ (not classified as a motion ideophone) is successively fruits, seeds, water, artefacts, small animals, an anaconda, a sloth and a sick human being. However, a number of motion ideophones are likely specific to animate figures, for example those expressing ‘walking’ and body-internal motion like ‘startling’.





ideophones with other semantics as well, and is often made clear by the valency of the verb in collocation.

To summarize, ideophones express events, as do verbs, but they do so in different ways, lexicalizing different semantic components from particular verbs. They can lexicalize information on the event itself, its manner, its spatial configuration, the participants, and the ground (see also Nuckolls (2021)).

#### **8.4. Phonosemantic characteristics of ideophones**

“Sound-symbolism is the direct linkage between sound and meaning” (Hinton et al. 1994). Phonosymbolism has been described as a property of at least some ideophones, consistent but not automatic and universal (Childs 1994: 194; Dingemanse et al. 2016; Nuckolls 1999). Some authors have tried to establish it as a rigid system where each vowel and each consonant could be given an iconic meaning, covering the whole range of ideophones (Langdon 1994; Egbokhare 2001). This endeavor does not seem realistic for Teko ideophones.<sup>35</sup>

Several methodological issues emerged when seeking to uncover and weigh any phonosemantic characteristics of Teko ideophones. The first was that it was not technically possible with the database available to compare ideophones with verbs and nouns, since they were not coded for semantics. Taking the stance that a study on phonosemantics focusing on ideophones alone should be sufficient, it then also proved difficult to weigh the frequency of association of a phonological feature and a semantic feature within the list of ideophones. The idea was then to compare the proportion of ideophones with both a particular phonological and a particular semantic feature to the number of ideophones with only one of those in the list, i.e. comparing for example how many ideophones start with /p/, how many depict an event involving the ‘hand’ and how many associate both features, to evaluate whether this association is significant. In all cases, a statistic analysis seemed unlikely to be fruitful, the semantic features being either coded on a small number of items, or not coded, because the list of coded semantic features was limited. Adding ad-hoc semantic features while carrying out a phonosemantic analysis would have been very subjective and circular. For these reasons, this section will only offer some preliminary insights into likely phonosemantic associations, at three levels: root

---

<sup>35</sup> A reviewer has constructively suggested that another way to approach phonosymbolism is the theory-driven and cross-linguistic approach, as found for example in the work by Johansson et al. (2020). It might be useful to consider how general theories about form-meaning associations may be reflected in Teko ideophones.

structure, vowels, and consonants. Maybe in the future, with a larger database and more coding, these could be better weighted.

### 8.4.1. Root structure

Recall that Teko ideophones are often monosyllabic, and generally start and end in a consonant, with the typical form (CV)<sup>n</sup>.CVC. Under this basic template, the longer the word, the richer its semantics. Also, the more unexpected the phonemes (such as a median consonant not being /r/ or identical with C1, or V2 being different from V1), the more complex the semantics. This raised complexity can concern the specification of aspect, participants, manner or ground for example (see the first four lines of Table 17). As for ideophones that depart from this canonical structure, a few observations are proposed (illustrated in the bottom part of the table). Long ideophones not made up of a sequence of CV syllables depict more complex events. Ideophones ending in a vowel generally refer to events without an intrinsic endpoint. And those with the CVV form almost all depict some motion in the air without the goal being part of their semantics.

root structure	ideophone	gloss
CVC	<i>kir</i>	grating
CV <sub>1</sub> rV <sub>1</sub> C	<i>kirig</i>	tying with the hand
CV <sub>1</sub> rV <sub>2</sub> C	<i>kirog</i>	swallowing voraciously
C <sub>1</sub> V <sub>1</sub> C <sub>2</sub> V <sub>1</sub> C <sub>1</sub> V <sub>1</sub> C <sub>2</sub> V <sub>1</sub> C	<i>kisikisig</i>	moving and scraping
CVVCVCVCVC	<i>paipipig</i>	stumbling
CV	<i>po</i>	leaving
CVV	<i>pai</i>	body falling

Table 17. Root structure of ideophones and phonosemantics.

### 8.4.2. Vowels

A number of studies on ideophones have shown that vowel substitution within the same consonant template can express variation in the general meaning of this abstract template (a very nice case study is Tufvesson 2011). Most of these focus on the quality of the vowel, and concord with the so-called Frequency Code (Hinton et al. 1994: 10) according to which vowels with high second formants /i/ are associated with small size, sharpness, fast movements, while vowels with low second formants like /u/ associate with large size, softness, heavy and slow movements.



Similar observations can be made with different templates. In a different semantic area, the contrast between *miŋ* ‘closing eyes’ and *moŋ/muŋ* ‘darkness’ can also be attributed to the quality of the vowel, the front one referring to a more delimited space. It has not yet been investigated whether the vowel quality has identical entailments in ideophones that do not appear to participate in a consonant template.

Additionally, special attention has been given to ideophones without monovocality, as they are formally marked. A preliminary study looked at ideophones with a back-front or front-back disharmony, with 8 and 10 items each. One could propose that these ideophones express some irregularity, or a fast change of state. However, the directionality of the change of vowel quality is not obviously significant. Both back-front and front-back combinations depict for example ‘opening eyes’ (*tuni, miwog*) and ‘swallowing’ (*sōʔēsōʔē, petog*).

### 8.4.3. Consonants

Table 18 lists a number of consonant-meaning associations that seem to be frequent in ideophones above chance, given the frequency within the list of ideophones of both the phoneme in that position and the meaning. Needless to say, these are frequent rather than universal associations.

consonant distribution	semantics	example	gloss of example
initial /s/	liquid	sokoɕ*	walking in mud
initial /h/	mouth	hĩ	showing teeth
initial /p/	hand	pog	hitting
initial /p/	air	pur	farting
initial /w/	soft ground	worog*	feeling a gap in the ground
initial nasal	vision	niŋ	winking
initial /k/	cutting and separating	kir	go off
initial voiceless plosive	contact by impact	pag	hitting
final velar /g/ or /ŋ/	ingestion	kĩrog	swallowing voraciously
final palatal /ɕ/ or /ɲ/	liquid or soft ground	soɕ*	entering water
final /g/	hand	kĩrig	tying with the hand
final /ŋ/	liquid or gaseous ground	mirimimiŋ*	liquid dripping softly

Table 18. Consonants in ideophones and phonosemantics.

I will simply illustrate some of these associations with two contrastive pairs. Examples (75) and (76) both end in /ɲ/ and evoke a liquid participant and some gaseous ground, and the initial consonant substitution fits the association of initial nasals with ideophones within the semantic domain of vision. As for (77) and (78), they both

depict a body motion with some body parts going down to the ground, but in (78) ending in /ɖʒ/ the ground is liquid.

(75) *tɥŋ*  
 ‘water pouring’

(76) *mɥŋ*  
 ‘invisible’

(77) *turug*  
 ‘stepping down’

(78) *turuɖʒ*  
 ‘fists hitting water’ (a local manifestation of pleasure)

## 9. Discussion

This paper has presented a comprehensive description of the ideophones of Teko, through a systematic comparison with nouns and verbs, at various levels of analysis. The first goal was to evaluate how marginal or central ideophones are within the Teko grammar, and the second goal was to determine whether and how they could be identified as forming a word class. Table 19 summarizes the findings per level of analysis (first column) with regard to whether and how ideophones are regular within the language (second column), or marginal (third column), and how they differ from the major lexical classes of nouns and verbs (fourth column).

	regular	marginal	different than nouns or verbs
frequency in the lexicon (1.3)	yes (5 to 8% of roots)		
frequency in speech (1.3)	yes (4% of words)		
phonemic inventory (2.2)	yes (29 phonemes in common)	yes (two sounds)	
phonotactics (2.3)	yes (same restrictions in general)	yes (final /h/ in two items)	different frequencies of distribution
nasality (2.4)	yes (mostly regular)	yes (a few irregularities)	

	regular	marginal	different than nouns or verbs
root structure (3.2)	yes (same syllable structures)	yes (two potential exceptions)	different distribution of structures
stress (4.2)		yes (some special stress patterns)	
prosody (4.3 to 4.5)		partially (salient mean vowel duration, pitch and intensity)	
expressive use of prosody (4.3 to 4.5)	yes (found with other word classes)		to be investigated
morphology (5.2)	yes (absence also found in other types of root)		no bound morphology
duplication (5.3)	yes (also on verbs)		different patterns than on verbs
word-class changing derivation (5.4)	yes		
syntax (6)	yes (integrated constructions)	yes (independent constructions)	different types of syntactic integration
sentence types (7.1)	yes		mostly in declarative sentences
discourse genre (7.2)	yes		
series (7.3)		yes (series of more than two items)	
semantics (8)	yes (comparable to that of verbs)	yes (depicts rather than refers)	
phonosemantics (8.4)		yes (some)	

**Table 19.** Integration of ideophones in the Teko language.

The detailed presentation of Teko ideophones developed in this paper and summarized in Table 19 shows that they are in fact rather well integrated in the lexicon, i.e. they are not outside of grammar, contrary to the outdated exotic view of ideophones (see also Newman 2001). Basically, only stress, prosody and their phonosemantics set them aside from the rest of the Teko lexicon. This paper also finds that the “exotism” of ideophones is limited. Particular facts in that direction listed in the third column are often restricted to a few items, i.e. a small minority of the ideophone class. Beside the stress patterns that definitively mark some of the ideophones as categorically different from the rest of the lexicon, the other major particularities of ideophones are not categorical: their prosody is discrete, and their phonosemantics is difficult to evaluate. The one major characteristic of ideophones that is fundamental to them is their depictive power. Depiction in ideophones has generally been approached through multimodal studies (Kita 1997; Reiter 2011; Dingemanse & Akita 2017). Unfortunately, the dataset used for this particular study does not allow us to study gestures, due to the absence of video recording. The marking of depiction in Teko ideophones has been investigated in detail from two of Akita (2021)’s perspectives: the foregrounding strategy (phonological distribution, root structure, word-level prosody) and the framing strategy (syntactic construction, pauses).

Finally, and most importantly for this special issue, this paper has also shown that Teko ideophones show regularities that call for them to be considered a distinct word category. Most of these features are statistical rather than categorical, except the absence of bound morphology, which seems to be their most straightforward formal defining criterium. My personal ordered list of tests used in methodologically identifying ideophones is as follows:

- i) identification of the word as a non-verb through absence of morphology
- ii) identification of the word as a non-noun on the basis of translation (either absence of translation in the text translation or “translation” with a full sentence, or an example (in elicitation), rather than by a simple word)
- iii) identification of the word as an ideophone if the item in question is found in constructions where ideophones are found (with their special position)
- iv) decision confirmed by expressive prosody when present (or if present in a different token of the item)
- v) confirmation by a consultant familiar with the metalinguistic terminology, if possible.



Returning to the central idea that ideophones depict rather than refer, this fact results from the accumulation of small or large differences with respect to the rest of the lexicon. This makes the ideophones in Teko an obvious category of the language, immediately accessible to both native and non-native speakers.

## Acknowledgments

I would first like to warmly thank students and trainees for their collaboration on the databases: Mao-Line Fillon, Kenza van den Heuvel, Yana Akvilina and Viviane Ribes. I address special thanks to Jennifer Krzonowski for precious help with Praat scripts and statistics. I would also like to thank Elsa Oréal and Jean-Luc Chevillard for coordinating a research project of the CNRS Fédération de Typologie et Universaux called *Vers une typologie des parties du discours mineures: onomatopées, idéophones, interjections* (2002-2005), as well as Aimée Lahaussais and Yvonne Treis for putting together a workshop on *Ideophones and interjections* at the Annual meeting of the SLE in 2019, later joined by Julie Marsault as editors of the current special issue. This paper has also benefited from the remarks and suggestions of two anonymous reviewers, which I sincerely thank here. Last but not least, I send my deepest gratitude to the Teko speakers who have allowed me to investigate their language two decades ago, and especially to Jammes Panapuy with his instant grasping of the concept of ideophones. I also want to thank Alexis Michaud for having shared his recordings with me long ago.

## Abbreviations

1INCL = 1st inclusive	FOC = focus	PRO = pronoun
ASSERT = assertive	FUT = future	PURP = purpose
CAUS = causative	IDEO = ideophone	RECP = reciprocal
CL = classifier	IMP = imperative	RED = reduplication
CONCES = concessive	INTER = interrogative	REF = referential
CONT = continuative	INTENS = intensive	REL = relativizer
CONTR = contrast	MID = middle	RELN = relational
DEM = demonstrative	NEG = negation	SG = singular
DESID = desiderative	PL = plural	TOP_SW = topic-switch
DIM = diminutive	PL.S = plural of subject	
EXCL = exclamative	POSTP = postposition	

## References

- Akita, Kimi. 2021. A typology of depiction marking: The prosody of Japanese ideophones and beyond. *Studies in Language* 45(4). 865–886. (doi:10.1075/sl.17029.aki)
- Akita, Kimi & Imai, Mutsumi & Saji, Noburo & Kantartzis, Katerina & Kita, Sotaro. 2013. Mimetic Vowel Harmony. In Frellesvig, Bjarke & Sells, Peter (eds.), *Japanese/Korean Linguistics*, vol. 20, 115–129. Stanford: CSLI Publications.
- Aplonova, Katya & Guitang, Guillaume & Nikitina, Tatiana. 2022. Pseudowords and ideophones in West African traditional narratives. Bucharest. (Presented at the SLE Annual Meeting, Bucharest)
- Association Solidarité Guyane. 2000. *Euro en Guyane. Langue amérindienne: Teko*. Direction des relations avec les Publics et de la Communication du Ministère de l’Economie, des Finances et de l’Industrie.
- Birk, D.B.W. 2015. *The MalakMalak language, Daly River (Western Arnhem Land)*. B-45. Pacific Linguistics. (PDF) (doi:10.15144/PL-B45) (<http://sealang.net/archives/pl/pdf/PL-B45.pdf>) (Accessed November 17, 2022.)
- Boersma, Paul & Weenink, David. 2023. *Praat: doing phonetics by computer [Computer program]. Version 6.3.11*. (<http://www.praat.org>)
- Cachine, Jean-Marc & Couchili, Lucie & Maurel, Didier & Monerville, Eddy & Panapuy, Jammes. 2020. *Kaleta teko ’awu panaĩtsĩ ’awu ipope ma’ě - dictionnaire teko - français*. (<https://corporan.huma-num.fr/Lexiques/dicoTeko.html>)
- Childs, G. 1994. African ideophones. In Hinton, Leanne & Nichols, Johanna & Ohala, John (eds.), *Sound symbolism*, vol. Chapter 13, 178–204. Cambridge University Press.
- Couchili, Ti’iwan & Maurel, Didier & Queixalós, Francesc. 2002. Classes de lexèmes en émérillon. *Amerindia* 26/27. 173–208.
- Cusic, David. 1981. *Verbal plurality and aspect*. Stanford University.
- Dhoorre, Cabdulqaadir Salaad & Tosco, Mauro. 1998. Somali Ideophones. *Journal of African Cultural Studies* 11(2). 125–156.
- Dingemanse, Mark. 2011. *The Meaning and Use of Ideophones in Siwu*. PhD dissertation. Doctoral dissertation, Radboud University, Nijmegen.
- Dingemanse, Mark. 2012. Advances in the Cross-Linguistic Study of Ideophones. *Language and Linguistics Compass* 6(10). 654–672.

- Dingemanse, Mark. 2017. Expressiveness and system integration: On the typology of ideophones, with special reference to Siwu. *STUF - Language Typology and Universals* 70(2). (10.1515/stuf-2017-0018)
- Dingemanse, Mark. 2019. “Ideophone” as a comparative concept. In Akita, Kimi & Pardeshi, Prashant (eds.), *Iconicity in Language and Literature*, vol. 16, 13–33. Amsterdam: John Benjamins Publishing Company. (doi:10.1075/ill.16.02din) (<https://benjamins.com/catalog/ill.16.02din>) (Accessed September 10, 2019)
- Dingemanse, Mark & Akita, Kimi. 2017. An inverse relation between expressiveness and grammatical integration: on the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics* 53(3). 501–532.
- Dingemanse, Mark & Schuerman, Will & Reinisch, Eva & Tufvesson, Sylvia & Mitterer, Holger. 2016. What sound symbolism can and cannot do: testing the iconicity of ideophones from five languages. *Language* 92(2). 117–133. (doi:10.1353/lan.2016.0034)
- Doke, C. 1935. *Bantu Linguistics Terminology*. London: Longmans.
- Egbokhare, Francis Oisaghaede. 2001. Phonosemantic correspondences in Emai attributive ideophones. In Voeltz, F. K. Erhard & Kilian-Hatz, Christa (eds.), *Ideophones* (Typological Studies in Language), vol. 44, 87–96. Amsterdam: John Benjamins Publishing Company. (doi:10.1075/tsl.44.08egb) (<https://benjamins.com/catalog/tsl.44.08egb>) (Accessed October 19, 2022.)
- ELAN. 2022. (Version 6.3) [Computer software]. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. (<https://archive.mpi.nl/tla/elan>)
- Gabas, Nilson Junior. 1999. *A Grammar of Karo, Tupi (Brazil)*. UCSB, Department of Linguistics.
- Gijn, Rik van. 2010. Middle voice and ideophones, a diachronic connection. *Studies in Language* 34(2). 273–297.
- Gordon, Matthew & Rose, Françoise. 2006. Emérillon stress: a phonetic and phonological study. *Anthropological Linguistics* 48(2). 132–168.
- Güldemann, Tom. 2008. *Quotative indexes in African languages. A synchronic and diachronic survey* (Empirical Approaches to Language Typology 34). Berlin: Mouton de Gruyter.
- Hinton, Leanne & Nichols, Johanna & Ohala, John. 1994. Introduction: sound-symbolic processes. In Hinton, Leanne & Nichols, Johanna & Ohala, John (eds.), *Sound symbolism*, 1–12. Cambridge: Cambridge University Press.

- Ibarretxe-Antuñano, Iraide. 2017. Basque ideophones from a typological perspective. *Canadian Journal of Linguistics/Revue canadienne de linguistique*. Cambridge University Press 62(2). 196–220. (doi:10.1017/cnj.2017.8)
- Ibarretxe-Antuñano, Iraide. 2019. Towards a semantic typological classification of motion ideophones: The motion semantic grid\*. In Akita, Kimi & Pardeshi, Prashant (eds.), *Iconicity in Language and Literature*, vol. 16, 137–166. Amsterdam: John Benjamins Publishing Company. (doi:10.1075/ill.16.07iba) (<https://benjamins.com/catalog/ill.16.07iba>) (Accessed November 21, 2022.)
- Idiatov, Dmitry. 2005. The exceptional morphology of Tura numerals and restrictors: Endoclitics, infixes and pseudowords. *Journal of African Languages and Linguistics* 26(1). (doi:10.1515/jall.2005.26.1.31) (<https://www.degruyter.com/document/doi/10.1515/jall.2005.26.1.31/html>) (Accessed November 14, 2022)
- Jacques, Guillaume. 2013. Ideophones in Japhug (Rgyalrong). *Anthropological Linguistics* 55(3). 256–287.
- Johansson, Niklas Erben & Anikin, Andrey & Carling, Gerd & Holmer, Arthur. 2020. The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology* 24(2). 253–310. (doi:10.1515/lingty-2020-2034)
- Kilian-Hatz, Christa. 2001. Universality and diversity: Ideophones from Baka and Kxoe. In Voeltz, F. K. Erhard & Kilian-Hatz, Christa (eds.), *Typological Studies in Language*, vol. 44, 155–163. Amsterdam: John Benjamins Publishing Company. (doi:10.1075/tsl.44.13kil) (<https://benjamins.com/catalog/tsl.44.13kil>) (Accessed July 18, 2023)
- Kita, Sotoro. 1997. Two-dimensional semantic analysis of Japanese mimetics. *Linguistics* 35. 379–415.
- Kunene, Daniel. 2001. Speaking the act. The ideophone as a linguistic rebel. In Voeltz, F. K. & Kilian-Hatz, C. (eds.), *Ideophones*, 183–191. Amsterdam/Philadelphia: John Benjamins.
- Langdon, Margaret. 1994. Noise words in Guaraní. In Hinton, Leanne & Nichols, Johanna & Ohala, John (eds.), *Sound symbolism*, vol. Chapter 7, 94–103. Cambridge University Press.
- Levin, Beth. 1993. *English verb classes and alternations: a preliminary investigation*. Chicago: University of Chicago Press.

- Matisoff, James A. 1975. Rhinoglottophilia: the Mysterious Connection between Nasality and Glottality. In Ferguson, Charles A. & Hyman, Larry M. & Ohala, John (eds.), *Nasàlfest*, 265–288. Stanford: Language Universals Project, Department of Linguistics, Stanford University.
- Maurel, Didier. 1998. *Eléments de grammaire émerillon*. Paris: A.E.A. (Chantiers Amerindia 1.23).
- Maurel, Didier. 2000. *Taila. 1596-1789: deux siècles de relations inter-communautaires teko-kaliña*. Université M.Bloch, Strasbourg II. (Mémoire de Maîtrise d’Ethnologie)
- Maurel, Ti’iwan. 1991. *Dzawapinim o’olam tamandua* (Krik Krak). Maripasoula: Association des Amis de l’Ecole élémentaire de Maripasoula.
- Maurel, Ti’iwan. 1993. *Wlakala n?wã’e dzowoma’?kom omba’e* (Krik Krak). Maripasoula: Association des Amis de l’Ecole élémentaire de Maripasoula.
- Mayer, Mercer. 1969. *Frog, where are you?* (Dial Books for Young Readers). New York: Penguin.
- McLean, Bonnie. 2021. Revising an implicational hierarchy for the meanings of ideophones, with special reference to Japonic. *Linguistic Typology*. De Gruyter Mouton 25(3). 507–549. (doi:10.1515/lingty-2020-2063)
- Mihas, Elena. 2012. Ideophones in Alto Perené (Arawak) from Eastern Peru. *Studies in Language* 36(2). 300–344. (doi:10.1075/sl.36.2.04mih)
- Navet, Eric. 1994. Introduction. In Couchili, Ti’iwan & Maurel, Didier (eds.), *Contes des Indiens émerillon*, 1–11. Paris: Conseil International de la Langue Française.
- Newman, Paul. 2001. Are ideophones really as weird and extra-systematic as linguists make them out to be? In Voeltz, F. K. & Kilian-Hatz, C. (eds.), *Ideophones*, 251–258. Amsterdam/Philadelphia: John Benjamins.
- Nuckolls, Janis B. 1996. *Sounds like life: sound-symbolic grammar, performance, and cognition in Pastaza Quechua* (Oxford Studies in Anthropological Linguistics 2). New York: Oxford University Press.
- Nuckolls, Janis B. 1999. The Case for Sound Symbolism. *Annual Review of Anthropology* 28. 225–252.
- Nuckolls, Janis B. 2021. Ideophones in Pastaza Quechua. *tsl.44.22nuc*. (<https://benjamins.com/catalog/tsl.44.22nuc>) (Accessed October 12, 2021)
- Reiter, Sabine. 2011. *Ideophones in Awetí*. Kiel. Doctoral dissertation, Christian Albrechts University.

- Renault-Lescure, Odile & Grenand, Françoise & Navet, Eric. 1987. *Contes amérindiens de Guyane* (Collection Fleuve et Flamme). Paris: Conseil International de la Langue Française.
- Rodrigues, Aryon. 1984. Relações internas na família linguística Tupí-Guaraní. *Revista de antropologia* 27–28. 33–53.
- Rose, Françoise. 2002. My hammock = I have a hammock. Possessed nouns constituting possessive clauses in Emérillon (Tupi-Guarani). In Cabral, Ana Suely & Rodrigues, Aryon (eds.), *Línguas Indígenas Brasileiras. Fonologia, Gramática e História. Atas do I Encontro Internacional do GTLI da ANPOLL*, vol. 1, 392–402. Belem, Brésil: CNPQ & Universidade federal do Para.
- Rose, Françoise. 2003. *Morphosyntaxe de l'émérillon. Une langue tupi-guarani de Guyane française*. Doctoral dissertation, Université Lumière Lyon II.
- Rose, Françoise. 2005. Reduplication in Tupi-Guarani languages: going into opposite directions. In Hurch, Bernhard (ed.), *Studies on Reduplication* (Empirical Approaches to Language Typology 28), 351–368. Berlin/New York: Mouton de Gruyter.
- Rose, Françoise. 2007. Action répétitive et action répétée : aspect et pluralité verbale dans la reduplication en émérillon. *Faits de Langues* 29 (La reduplication). 125–143. (Ed. Bibliothèque Faits de Langues)
- Rose, Françoise. 2008. A typological overview of Emerillon, a Tupi-Guarani language from French Guiana. *Linguistic Typology* 12(3). 431–460.
- Rose, Françoise. 2009. A hierarchical indexation system: the example of Emerillon (Teko). In Epps, Patience & Arkhipov, Alexander (eds.), *New Challenges in Typology. Transcending the Borders and Refining the Distinctions* (Trends in Linguistics), 63–83. Berlin: Mouton de Gruyter.
- Rose, Françoise. 2011. *Grammaire de l'émérillon teko, une langue tupi-guarani de Guyane française* (Langues et Sociétés d'Amérique Traditionnelle 10). Louvain: Peeters.
- Rose, Françoise. 2018. *Teko lexical and textual Toolbox database*.
- Rose, Françoise. 2023. Word classes in Mawetí-Guaraní languages. *The Oxford Handbook of Word Classes* (Oxford Handbooks). Oxford University Press.
- Rose, Françoise & Vanhove, Martine. 2007. Discours rapporté direct et prosodie en émérillon et en bedja. Nantes. (Presented at the Colloque CerLiCO Grammaire et Prosodie, Nantes)

- Samarin, William. 2001. Testing hypotheses about African ideophones. In Voeltz, F. K. & Kilian-Hatz, C. (eds.), *Ideophones*, 321–337. Amsterdam/Philadelphia: John Benjamins.
- Smoll, Laetitia I. 2014. *Me:RuRu, Foku, and tSitowiS: An analysis of ideophones in Katuena (Tunayana)*. Master Thesis, Universiteit Leiden.
- Talmy, Leonard. 2000. *Toward a cognitive semantics*. Cambridge: MIT Press.
- Tufvesson, Sylvia. 2011. Analogy-making in the Semai Sensory World. *The Senses and Society* 6(1). 86–95. (doi:10.2752/174589311X12893982233876)
- van der Hulst, Harry & van de Weijer, Jeroen. 2011. Vowel harmony. In Goldsmith, John A. (ed.), *The Handbook of Phonological Theory*, 495–534. John Wiley & Sons.

**CONTACT**

francoise.rose@univ-lyon2.fr

# Ideophones in Kambaata (Cushitic): Grammar, meaning and use

YVONNE TREIS

CNRS, UMR 8135 LANGAGE, LANGUES ET CULTURES D'AFRIQUE (LLACAN)

Submitted: 31/01/2023    Revised version: 23/08/2023

Accepted: 04/11/2023    Published: 12/06/2024



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

## Abstract

In the literature on Cushitic languages, ideophones have often only been treated in a cursory manner. A little explored problem of the synchronic analysis concerns their word class status: do they constitute a word class on their own, or should they be subsumed under another word class and if so, under which one? The study of Kambaata ideophones presented here shows that there are valid arguments both for analyzing them as a separate word class and as a subclass of verbs. Based on a language-internal definition of the ideophone word class, features of its phonology, phonotactics and stress marking are discussed. The section on morphosyntax shows in which syntactic functions ideophones are used, clarifies the status of their support verbs ('say' and 'do') and considers the argument structure of ideophones. The section on ideophone-related word formation investigates which derivational processes take ideophonic inputs, describes the functions of reduplication and illustrates the significance of compounding and conversion for the formation of ideophones. The discussion of the semantics of ideophones is followed by preliminary information about the frequency of ideophones across text types. Finally, the use of 'say' as a quotative verb in reported speech constructions is explored, and it becomes clear that the line between ideophone constructions with 'say' and reported interjections is not always easy to draw.

**Keywords:** Cushitic; ideophone; interjection; word class; stress.



## 1. Introduction

Kambaata (Highland East Cushitic, ktb; Glottolog code: kamb1316; endonym: *Kambaatissata*) is a morphologically rich language, which requires all nouns, pronouns, verbs and adjectives to be inflected (Treis 2023a). It therefore catches the eye that two major word classes are morphologically invariant: ideophones, e.g. *xóqq* ‘rise’, in (1) in combination with the verb *y-* ‘say’, and interjections, e.g. *hashshú* ‘yay, great, congratulations’ and *ekku* ‘okay’ in (2148).

- (1) *Mexx-é*            *afuu'll-itóo*                            *ma'nn-éechch*    *xóqq = y-itán*  
 single-MULT    sit\_down-3F.PFV.REL            place-F.ABL    rise.IDEO = say-3F.PFV.CVB  
*qax-ée*            *wól-i-s*                            *óos-ut*                            *me'-ú = rr-a-ndo*  
 extent-M.DAT other-F.NOM-DEF children-F.NOM how\_many-M.ACC = NMZ4-M.ACC-Q  
*hujat-táa'u*  
 work-3F.IPFV  
 ‘How many (tasks) do the other children work before she stands up once from the place where she is sitting?’ (Field notes 2015, DWD)

- (2) *Hashshú,*    *góon-ch-u-ta,*                            *abb-íshsh*  
 yay.INTJ    males-SGV-F.PRED-F.COP2    exceed-CAUS1.1SG.PFV.CVB  
*galaxx-áan-ke.*  
 thank-1SG.IPFV-2SG.OBJ  
*Y-itoontí = r-a*                            *gu'mm-á <n> ka*    *ekku!*  
 say-2SG.PFV.REL = NMZ4-M.ACC    all-M.ACC <EMP>    okay.INTJ  
 ‘Yay, you are a hero! I thank you very much. Okay to all what you said.’ (SMS exchange 2019, BN)

In the literature on Cushitic languages, ideophones (as well as interjections) have so far only been treated in a cursory manner and their importance has (potentially) been underestimated. Tosco hypothesizes that ideophones “are not so common in Gawwada (gwd) (nor generally in Cushitic) as they are in many other African languages” (2006: 887). Mous’s detailed overview chapter of Cushitic dedicates only half a page to ideophones (2012: 381–382); interjections are not discussed at all. To

date, there are only few detailed studies on ideophones in individual languages, the most notable exceptions are Cabdulqaadir & Tosco (1998) on Somali (som) and Tosco (2006) on Gawwada, both languages of study belonging to the Lowland East Cushitic branch. In addition, three qualifying works from Addis Ababa University deal with ideophones in Cushitic languages (Amina 2013; Girum 2013; Desalegn 2020). While ideophones are underrepresented in descriptive and documentary works, so-called “descriptive compounds”, i.e. complex predicates whose semantic core is a non-inflecting coverb (e.g. an ideophone) followed by an inflecting, semantically bleached light verb (often ‘say’ and ‘do’), have attracted the attention of areal linguists, who trace their distribution across Northeastern Africa, and historical linguists, who see them as the source construction for new verb inflections across Afroasiatic (see, for instance, Cohen et al. 2002; Güldemann 2005).

An important, little explored problem of the synchronic analysis of ideophones in Cushitic concerns their word class status: do they constitute a word class on their own, or can they be subsumed under another word class and if so, which one? For Tosco, ideophones in Cushitic languages are “undoubtedly [...] nominal in character” (2006: 887; see also Cabdulqaadir & Tosco 1998 on the nominal nature of Somali ideophones). The detailed study of Kambaata ideophones presented here significantly expands an earlier study of onomatopoeic ideophones (Treis & Deginet 2024). The present paper argues that ideophones could either be analyzed as a separate open word class or as a subclass of verbs. After a general introduction into the grammar of the language (§2), the core of this paper (§3) aims at characterizing ideophones from different angles. First, the defining criteria for the word class are established (§3.1), then features of its phonology, phonotactics and stress marking are discussed (§3.2). The section on morphosyntax (§3.3) shows in which syntactic functions ideophones are used, clarifies the status of their support verbs (enclitic or independent) and their variant forms, and looks at the argument structure of ideophones. The next section (§3.4) delves into ideophone-related word formation, investigates which derivational processes take ideophonic inputs, shows for which functions ideophones are morphologically reduplicated and illustrates the significance of compounding and conversion for the formation of ideophones. §3.5

is dedicated to the semantics of ideophones, while §3.6 provides preliminary information on the frequency of ideophones across text types and compares the size of the word classes of verbs and ideophones. §4 looks at the grammar of reported speech and other uses of ‘say’. The section highlights, among others, that it is not always possible to draw a line between ideophones and reported interjections. The final discussion in §5 weighs up the pros and cons of considering ideophones as a word class on a par with the major word classes of nouns, adjectives and verbs, or of subsuming verbs and ideophones as subclasses under a joint word class. The final section also addresses to what extent the Kambaata ideophones fit the comparative concept of ideophones proposed by Mark Dingemans.

The description presented here is based on data from my field notes<sup>1</sup> and transcribed recordings, local Kambaata publications and the extensive lexical documentation in Alemu’s (2016) Kambaata-Amharic-English dictionary. If not indicated otherwise, all ideophones presented in this paper are attested in more than one source. Data taken from Alemu (2016) has all been checked with a native speaker and the translations corrected or refined.<sup>2</sup> Recorded field data is marked by the initials of the recorded speaker, the date of recording and the file number (e.g. AN2016-02-19\_001). All data cited from local Kambaata publications has been segmented, stress-marked, glossed and translated by me. Wherever possible and relevant, complementary information about ideophones in related languages is given.

## **2. Background on Kambaata**

Kambaata is a Highland East Cushitic language spoken in southwestern Ethiopia by at least 600,000 speakers, according to the latest census (Central Statistical Agency

---

<sup>1</sup> The field notes 2021-2023 were collected remotely or in person with native speakers in Germany.

<sup>2</sup> Desalegn (2020) is a thesis on Kambaata ideophones, which is partly based on Alemu (2016). I was not (yet) able to confirm many of the new ideophones provided in the appendix (2020: 135–138), which is possibly related to the fact that the author and most of his consultants come from the Shiinshicho area, while my consultants are from around Duuraame, Daambooyya and Hangacca.

2007: 74). Neighboring groups are speakers of closely related Cushitic languages (Hadiyya; Alaaba)<sup>3</sup> and of Omotic languages (Wolaitta; Dawro).<sup>4</sup> The Ethiopian lingua franca Amharic (Semitic, amh) is the most important second language of Kambaata speakers. Kambaata's official orthography is based on the Roman script and used with minimal modifications to transcribe the data in this article. The following graphemes are not in accordance with IPA conventions: <ph> /p'/, <x> /t'/, <q> /k'/, <j> /dʒ/, <c> /tʃ/, <ch> /tʃ/, <sh> /ʃ/, <y> /j/ and <'> /ʔ/. Geminate consonants and long vowels are marked by doubling, e.g. <shsh> /ʃ:/ and <ee> /e:/. Consonant clusters consisting of a glottal stop and a simplex sonorant are spelled as trigraphs, e.g. <'rr> /ʔr/, to distinguish them from laryngealized sonorants, e.g. <'r> /r'/. Nasalization is marked by a macron, e.g. <ā> /ã/. The minor adaptations to the official orthography made in this contribution concern the additional marking of phonemic stress by an acute accent and the consistent notation of the phonemic glottal stop whenever it occurs in word-medial and word-final position.

Kambaata is a suffixing, agglutinating-fusional language with many portmanteau morphemes. It is both head- and dependent-marking with nine nominal cases and subject indexing on verbs. The constituent order is head-final: dependent clauses precede main clauses, the main verb or copula is the last constituent in a clause; in the NP, all modifiers (including relative clauses) are placed before the head noun.

The word classes can be divided into those that have morphologically variant (i.e. inflecting) members and those that have morphologically invariant members. Nouns inflect obligatorily for case and gender. The different types of pronouns are obligatorily marked for case as well as person, gender, number and/or honorificity. Adjectives must be minimally marked for case and gender; as modifiers, they agree with the head noun. All verb forms apart from the verbal noun are subject-indexed and marked for at least one other inflectional category (aspect, mood, subordination); verbal nouns are marked for case and gender. The non-inflected

---

<sup>3</sup> Hadiyya: hdy; Alaaba: alw.

<sup>4</sup> Wolaitta: wal; Dawro: dwr.

nominal, adjectival or verbal root is bound and never uttered in isolation. Pronominal roots are often not isolatable due to fusion or suppletion.

Of the five word classes with morphologically invariant members, three are very small. The word class of conjunctions contains only two members: *té* ‘or’ for the disjunction of noun phrases and *bagáan* ‘but’, a contrastive clause conjunction. Instead, Kambaata makes use of subordinating and coordinating suffixes. There is also a small word class of discourse particles with members such as *ammóo* ‘however, furthermore’ (contrast, addition), *áchche* ‘then’ (consequence), *áda* ‘oh!, I see’ (surprise, understanding) and *éman* ‘congratulations!’.<sup>5</sup> The third negligibly small word class are adverbs, encompassing *léelan* ‘slowly’, *dángo* ‘unexpectedly’ and *bíinin* ‘separately’. In the place of adverbs, Kambaata uses nouns and adjectives marked for adverbial cases or converbs in adverbial function. The two remaining invariant word classes, ideophones and interjections, have a large number of members. Ideophones are at the core of this paper, whereas features of interjections are here treated only insofar as they are relevant for the description of ideophones (in §4).

### 3. Features of ideophones

#### 3.1. Language-internal definition and terminological choices

Ideophones are defined in Kambaata as an open word class of morphologically invariant lexemes that require the support verbs *y-* ‘say’ in intransitive clauses and *a’-* ‘do’ in transitive clauses to be inflected and syntactically integrated into an utterance. Ideophones differ from the equally invariant interjections in that they do not constitute an utterance on their own. In rare contexts, which are elliptic and belong to the domain of artistic speech, ideophones are attested in independent use. This is, for instance, seen in a riddle from Desalegn (2020: 111), here presented in

---

<sup>5</sup> The list of discourse particles is possibly incomplete. For the use of *éman* and *áda*, see (23) and (53).

(3).<sup>6</sup> See also verse 7 in the song *Haate Haate Haate* in Treis (2012), which ends in an ideophone without a support verb.

- (3) *Cuph-á*            **dúbb**,            *can-á*            **wíqq**  
 pestle-M.ACC    pound.IDEO      leather\_bag-M.ACC    tie\_tightly.IDEO  
 (Riddle) ‘A pestle – to pound, a leather bag (filled with air) – to tie up tightly.’  
 (Desalegn 2020: 111; transcription, glossing and translation adapted; solution  
 of the riddle provided by YYZ: *cinú* ‘excrements’)

Ideophones principally combine with two different support verbs (4)-(5) (see §3.3.4 for exceptions).<sup>7</sup> The support verb *y-* ‘say’ is used in intransitive contexts, the support verb *a’-* ‘do’ in transitive contexts. As ideophone support, *y-* ‘say’ varies freely with *ih-* ‘become’, and *a’-* ‘do’ with *ass-* ‘do’ (§3.3.3); the alternatives are, however, less commonly used. As argued in §3.3.2, the support verbs are enclitic to the ideophones. All support verbs are still used as regular verbs in isolation, where they have retained their full lexical meaning (§4). The full reduplication of ideophones, as in (5), is frequent (§3.4.3).

- (4) *bínn* = *y-* ‘be dispersed, be scattered’  
*bínn* = *a’-* ‘disperse (s.th.), scatter (s.th.)’ – see (41) for its use in an example
- (5) *dúbb* = *y-* / *dúbb~dúbb* = *y-* ‘produce the *dúbb*(~*dúbb*) sound (of feet, flail, load hitting the ground, pestle hitting the wooden mortar), be pounded’  
*dúbb* = *a’-* / *dúbb~dúbb* = *a’-* ‘cause to produce the *dúbb*(~*dúbb*) sound; pound; thresh’

The term “support verb” used here is equivalent to the more common term “light verb” found in the literature on complex predicates. The construction in which a

<sup>6</sup> Girum (2013: 52–53) also reports about the independent use of ideophones in Sidaama (sid) riddles.

<sup>7</sup> For reasons of space, in all the data tables below, ideophones are presented with only one possible support verb, often the one that is most commonly attested in my database. However, if not indicated otherwise (as in §3.3.4), these ideophones can usually also combine with the transitive or intransitive counterpart.

non-inflecting ideophone (or another coverb) is combined with a support verb is referred to in the literature on languages of Northeast Africa by a variety of terms, among them “compound verb”, “composite verb” and “descriptive compound”,<sup>8</sup> and considered an areal feature of the languages of the Ethiopian Linguistic Area or Northeast Africa (Leslau 1945: 72; Ferguson 1976: 71–72; Zaborski 1991: 125; Appleyard 2001: 1–2; Güldemann 2005). For Kambaata, I first speak of “ideophone-support constructions”, before arguing in §5 that “periphrastic verb” could be a possible alternative.

### 3.2. The phonology, phonotactics and stress marking of ideophones

When compared to lexemes of other word classes, ideophonic lexemes do not have a particularly unusual phonology. The vast majority are phonologically inconspicuous. Only a handful contain phonemes that are elsewhere marginal (Treis & Deginet 2024), such as nasal vowels, e.g. *hāā=y-* ‘gape (of mouth, abyss, deep hole)’ and the geminate glottal stop, e.g. *mé”a=y-* ‘bleat (of goat)’. The geminate glottal stop is, however, also found in interjections, e.g. *há”a* ‘yuck’, and the nasal vowels are attested in the noun *hāy-í* ‘front leg of cattle’ and in several interjections, e.g. *í* ‘take what I have in my hand’ and *āā* ‘yes’. The onomatopoeic ideophones *nyáu=y-* ‘meow (of cat)’, *wúh=y-* ‘bark (of dog)’ and *buhhú=y-* ‘cough, make a coughing sound’ are the only lexemes known to contain a palatal nasal,<sup>9</sup> a word-final glottal fricative and a geminate glottal fricative, respectively. At first view, the frequency of certain phonemes in ideophones is not significantly different from that in lexemes of other word classes, but this question needs to be studied quantitatively.

---

<sup>8</sup> Some languages in Northeast Africa have productive processes by which non-inflecting forms are derived from verbal lexemes. These constructions with a verb-based coverb are pragmatically marked when compared to simple verbs; see the “intersubjective modal functions” reported for Afar (aar) by Cohen et al. (2002: 247) or the attenuating/intensifying function of deverbal coverbs in Amharic (Meyer this volume). In this context, “descriptive” serves as a cover term for “pragmatically marked”.

<sup>9</sup> Alemu (2016: 755) transcribes ‘meow’ with an alveolar nasal as *naawu*.

Regarding phonotactics, the typical Kambaata ideophone is monosyllabic, less commonly disyllabic, and ends in a geminate consonant or consonant cluster;<sup>10</sup> see selected examples in Table 1.<sup>11</sup>

Ideophone	English
<b>Monosyllabic in CC#</b>	
<i>kú'nn = a'</i>	'tip (s.th.) out'
<i>qóss = y-</i>	'squat, sit on one's heels, sit down (of dog, cat)'
<i>shóott = y-</i>	'get up abruptly'
<i>wíll = y-</i>	'sneak away, disappear (from s.o.'s view)'
<i>xééphph = y-</i>	'be about to do'
<i>zágg = y-</i>	'fall flat to the ground (e.g. after heavy beating, when fainting)'
<b>Disyllabic in CC#</b>	
<i>canqárr = y-</i>	'cling (to s.o./s.th.), hold on (to s.o./s.th.)'
<i>fitákk = a'</i>	'untie (s.th.)'
<i>fokkótt = y-</i>	'bow (down), take a bow'
<i>habább = y-</i>	'burn (vi) with a sudden and high flame (e.g. of dry leaves)'
<i>hokkóbb = y-</i>	'stand on the hind legs and lean the front legs against (e.g. a tree)'
<i>qomfó'll = y-</i>	'get a dent, be dented'

**Table 1:** Ideophones illustrating the most common phonotactic structures.

Other phonotactic structures are not excluded, albeit less common (Table 2).<sup>12</sup> Some ideophones end in a single consonant or a long or short vowel. If an ideophone is vowel-final, it is more likely to be disyllabic than monosyllabic.

<sup>10</sup> Crass (2005: 226–227) reports the same phonotactic preference for K'abeena (alw) ideophones.

<sup>11</sup> Recall from §2 that certain consonants are represented by digraphs in the official orthography, e.g. <ph> /p'/ and <phph> /p':/, and that glottal stop-sonorant clusters are represented by trigraphs, e.g. <'nn> /ʔn/. The enclitic support verbs in this and the following tables are *y-* 'say' and *a'* 'do'.

<sup>12</sup> Only tendencies can here be described. Due to the large number of ideophones attested in the Kambaata sources (§3.6), many hundreds still need to be checked in the field.



Ideophone	English
<b>Monosyllabic in C#</b>	
<i>fúq</i> = y-	‘gush out, spill out’
<i>húf</i> = y-	‘boil over (of liquid in cooking pot), rise through fermentation’
<i>láp</i> = y-	‘lie down a bit (e.g. of people taking a nap)’
<i>shíq</i> = y-	‘move over, scoot over’
<i>túk</i> = y-	‘walk behind each other in a large group’
<b>Monosyllabic in V(V)#</b>	
<i>tú</i> = y-	‘spit’
<i>fúu</i> = a’-	‘blow (of strong wind)’
<i>táa</i> = a’-	‘shoot (a gun)’
<i>wóo</i> = y-	‘wail’
<i>úu</i> = y-	‘say mm-hmm, give sign of attentive listening, backchannel’
<b>Disyllabic in C#</b>	
<i>chalál</i> = y-	‘float’
<i>hagág</i> = y-	‘strut about’
<i>hambúq</i> = a’-	‘down (a drink), finish (a drink) in one go’
<i>shigíg</i> = a’-	‘shock, disgust, revolt (s.o.)’
<i>xambáq</i> = a’-	‘stick, attach (s.th.)’
<b>Disyllabic in V(V)#</b>	
<i>culú</i> = a’-	‘please (s.o.), seem beautiful (to s.o.)’
<i>forgó</i> = y-	‘move a bit away, keep one’s distance’
<i>hamúu</i> = a’-	‘drive (s.o.) insane, make (s.o.) lose their mind’
<i>kachá</i> = y- <sup>13</sup>	‘argue, thrash out a problem’
<i>rejé</i> = y-	‘become calm’

**Table 2:** Ideophones illustrating less common phonotactic structures.

Trisyllabic ideophones (Table 3) are considerably less frequent than disyllabic ones.<sup>14</sup>

<sup>13</sup> This ideophone was incorrectly given as an example with penultimate stress in Treis (2008: 87).

<sup>14</sup> Desalegn (2020: 55–56) reports about simplex ideophones with more than three syllables. However, in the examples given, the support verb *a’*- ‘do’ is erroneously included in the syllable count, e.g. *munxulúq* = a’- ‘remove completely’ (*munt’úluk’k’a?* in his writing) is considered to have four syllables.

Ideophone	English
<i>biciríqq = y-</i>	‘make a U-turn’
<i>bitikínn = a’-</i>	‘discard (e.g. an idea)’
<i>budulúmm = y-</i>	‘struggle, try hard’
<i>buxurúqq = y-</i>	‘come out suddenly, pop out’
<i>chachachá = y-</i>	‘be sufficiently fermented (of beer)’
<i>kushurúmm = a’-</i>	‘crunch, eat (s.th. hard) noisily’
<i>menxeléqq = a’-</i>	‘open completely/widely, uncover, disclose’
<i>munxulúqq = a’-</i>	‘remove completely’
<i>ororó = a’-</i>	‘soothe (a crying child) with a lullaby and rocking’
<i>tililí = y-</i>	‘ring (of telephone)’

Table 3: Trisyllabic ideophones.

In the vast majority of cases, the vowels across multisyllabic ideophones are identical; see most examples in the lower halves of Table 1 and 2 as well as in Table 3. Among trisyllabic ideophones, exceptions to this rule are especially hard to come by; but see *haburútt = y-* ‘wilt (of plants); become meagre, get a gaunt face (due to disease)’.

Phonotactically, ideophones differ little (if at all) from other roots in the language. Most verb roots, for instance, also have one or two vowels as nucleus (e.g. *laq-* ‘head (towards a place)’, *zuug-* ‘scrape; tan’, *íill-* ‘arrive’; *dagud-* ‘run’, *fikkaan-* ‘have many descendants’, *sangaagg-* ‘lean back one’s head’), while roots with three vowel nuclei (e.g. *foroffiit-* ‘steam’) are infrequent. Verbs roots also tend to have identical vowel nuclei if they are multisyllabic; this tendency seems, however, less pronounced than in the case of ideophones.

Kambaata is a language with phonemic stress whose position in a word is predominantly grammatically determined. The bound verbal, adjectival and (pro)nominal roots are undetermined for stress, but all inflectional morphemes are realized segmentally by a suffix and suprasegmentally by a specific stress position. Thus the stress patterns of word forms of the inflecting word classes are determined by their inflectional morphology. Kambaata has many systematic stress minimal pairs; see, for instance, the difference between accusative *-ú* vs. nominative *’-u* in nouns of the masculine declension M3: *bóos-u* (M.NOM) vs. *boos-ú* (M.ACC) ‘large clay pot’. In contrast, in the non-inflecting word classes of ideophones and interjections, the position of the stressed syllable is lexically determined and unpredictable, i.e. it

has to be noted in the dictionary.<sup>15</sup> Overall, ultimate stress is far more frequent on ideophones (see most examples in this article) than penultimate stress (Table 4). Trisyllabic ideophones with antepenultimate stress (<sup>l</sup>σσσ) are not attested.

Ideophone	English
<i>báa'a=y-</i>	'bleat (of sheep)'
<i>bíira=y-</i>	'forgive (esp. in a reconciliation ceremony)'
<i>cá'a=a'-</i>	'clear, remove, clean completely'
<i>dú''u-dú''u=y-</i>	'beat (of heart)'
<i>háā''ā=y-</i>	'hee-haw, bray (of donkey)'
<i>hánda=y-</i>	'become thankful, grateful, content; become gratifying'
<i>hó''a=a'-</i>	'cause to vomit'
<i>ilíi=y-</i>	'ululate'
<i>jába=y-</i>	'bless (before pouring coffee)'
<i>mé''a=y-</i>	'bleat (of goat)'
<i>qúuxo=y-</i>	'do deliberately, willfully, on purpose, according to a plan'

Table 4: Ideophones with penultimate stress ((σ)<sup>l</sup>σσ).

Every simplex ideophone has one prominent (stressed) syllable. Stress thus helps distinguish between ideophones that have two (or more) identical syllables, i.e. which are lexically reduplicative, from ideophones that are morphologically reduplicated. The former have one stressed syllable, e.g. *tatá=y-* 'fall suddenly and heavily (of rain)', while the latter are stressed on each reduplicant, e.g. *cíl~cíl=y-* 'become shiny (with oil, lotion)' and *káf~káf=y-* 'put oneself in danger'. For information on the stress marking of the support verbs, see §3.3.3.

Ideophones display a higher degree of phonological variation across speakers than lexemes of other word classes (a similar observation is made for Sidaama ideophones in Girum 2013: 34). Variation affects vowel quality (e.g. *CaCaC ~ CuCuC*), the mode of articulation of consonants (e.g. *ph /p'/ ~ b, q /k'/ ~ g, c /tʃ'/ ~ sh /ʃ'/*) and even syllable structure (e.g. *CVCC ~ CVCCí*), as a perusal of the available sources and my experience in work with different speakers show; see Table 5 for illustrative examples. Deciding which form of an ideophone to provide

<sup>15</sup> Unfortunately, the stress of ideophones and interjections is not marked in Alemu (2016). Desalegn (2020: 61) erroneously states that ideophones are accented on the "second syllable"; irrespective of whether this is to be interpreted as second syllable from the left or right, the statement is falsified by my data and even by the data that he himself presents in the thesis.

in a wordlist is, therefore, considerably more challenging than for lexemes of other word classes.

Variant 1	Variant 2	English	Sources
<i>bárr = y-</i>	<i>búrr = y-</i>	‘fly’	V1: TD2016-02-11_001 / V2: assistant <sup>16</sup>
<i>birxí~birxí = y-</i>	<i>bírx~bírx = y-</i>	‘twist, wriggle about (e.g. to free oneself)’	V1: Frog Story DWD, Kambaata Education Bureau (1989: 7.125) / V2: Alemu (2016: 136)
<i>bolóng~bolóng = a’</i> -	<i>bolónq~bolónq = a’-</i>	‘move one’s eyes (up/down, left/right)’	V1: TD2016-02-11_001, see 160 / V2: assistant
<i>chílk~chílk = y-</i>	<i>chilkí~chilkí = y-</i> <i>kilchí~kilchí = y-</i>	‘clink (of keys, coins, small bells on a horse’s neck)’	V1 + V2 + V3: Treis & Deginet (2024)
<i>hokkóbb = y-</i>	<i>hochhóbb = y-</i>	‘stand on the hind legs and lean the front legs against s.th.’	V1 + V2: Alemu (2016: 493)
<i>kirír = a’-</i>	<i>kurúr = a’-</i>	‘rotate’	V1 + V2: Alemu (2016: 596)
<i>kuukú = y-</i>	<i>guugú = y-</i>	‘coo (of dove)’	Treis & Deginet (2024)
<i>lácc = y-</i>	<i>láshsh = y-</i>	‘go slowly, slow down (vi)’	V1 + V2: field notes and different written sources
<i>ríph~ríph = y-</i>	<i>ríbb~ríbb = y-</i>	‘move slightly back and forth, up and down, in waves’	V1: Kambaata Education Bureau (1989: 4.51) / V2: Alemu (2016: 870)
<i>qacác = a’-</i>	<i>qucúc = a’-</i>	‘crunch (e.g. roasted grain, bones)’	V1 + V2: Alemu (2016: 789, 851)
<i>qu’mmí = eecc-</i>	<i>quí’m = eecc-</i>	‘bring together (for one’s benefit)’	V1: AN2016-02-19_001, see 168 / V2: assistant
<i>qúrc = a’-</i>	<i>gúrc = a’-</i>	‘swallow’	V1 + V2: Alemu (2016: 404)
<i>shóott = y-</i>	<i>shokótt = y-</i>	‘get up abruptly’	V1 + V2: Kambaata Education Bureau (1989: 8.22; 6.125)

**Table 5:** Illustrative examples of variation in the pronunciation of ideophones  
(V1/V2 = pronunciation variants).

<sup>16</sup> “Assistant” stands for the native speaker who assisted in the transcription of a recording and who pointed out discrepancies between their and the recorded speaker’s pronunciation.

### 3.3. The morphosyntax of ideophones

#### 3.3.1. Syntactic functions of ideophones

With the support of ‘say’ and ‘do’, ideophones can be used in any syntactic context where verbs are used in Kambaata.<sup>17</sup> In (6), the last ideophone is a declarative main clause verb, while the preceding ideophones are converbs. In (7), the ideophone serves as a non-declarative, benedictive main clause verb.

- (6) **Míkk.míll**=y-itu’nnáachch      **címm**=y-ít  
 budge.IDEO = say-3F.NEG4.CVB    shrink.IDEO = say-3F.PFV.CVB  
*afuu’ll-ít,*                      *ill-í-se*                      *al-í,*                      *muggeenn-á,*  
 sit\_down-3F.PFV.CVB    eye-F.ACC-3F.POSS    up-M.ACC    down-M.ACC  
*gur-á[ta],*                      *makk-íta*                      **bolóng~bolóng**=at-táa’,  
 left-F.ACC                      right-F.ACC    RED~move\_one’s\_eyes.IDEO = do-3F.IPFV  
*tah-íchch-u*                      *waal-anó* = g-a                      *zug-gán*  
 flies-SGV-M.NOM    come-3F.IPFV.REL = SIM-M.ACC    lie\_in\_ambush-3F.PFV.CVB  
 (Speaking about a chameleon:) ‘It (lit. she) sat (there) stock-still, shrunk, (and only) moved its eyes up, down, to the left, to the right, (while) lying in ambush for a fly.’ (TD2016-02-11\_001)

- (7) *Maaráam-it*    *xumm-ûn*                      *gulub-í-kk*  
 Mary-F.NOM    peace-M.ICP                      knee-F.ACC-2SG.POSS  
**címm** = át-to-he  
 bring\_together.IDEO = do-3F.BDV-2SG.OBJ  
 (Blessing for a woman who has just given birth:) ‘May Mary bring your knees (i.e. here: pelvis) together (again) in peace!’ (EK2016-02-23\_002)

Ideophones can also be used as final constituents in relative clauses and relative-based adverbial clauses (8). In (9), the support verb enables the ideophone to be used as a case-inflected verbal noun complement of *hoog*- ‘not do’.

<sup>17</sup> There is one known exception to this rule: the ideophone *qúuxo*=y- ‘do deliberately, willfully, on purpose, according to a plan’ is only used in adverbial function and therein inflected as a converb.

- (8) (...) *resh-á*                    *xóqq = a'-eemá = g-a < n > ka*  
corpse-M.ACC    rise.IDEO = do-3HON.PFV.REL = SIM-M.ACC < EMP >  
*iddir-í*                                *bun-á*                                *azzaz-eenáyyoomma*  
funeral\_association-M.GEN    coffee-M.ACC    order-3HON.PROG  
‘(...) as soon as the corpse has been picked up, one orders the funeral association’s coffee.’ (EK2016-02-23\_003)

- (9) *Giir-áta*    *danáam-o = gga*                    *tú'mm = a'-ú*  
fire-F.ACC    good-M.OBL = SIM-M.OBL    narrow\_down.IDEO = do-M.ACC  
*hóog-gi-yan*                    *mát-oa = rr-áan*                    *aphph-íti-yan*  
not\_do-2SG.PFV.CVB-DS    one-M.OBL = NMZ4-M.LOC    seize.MID-3F.PFV.CVB-DS  
*bu < m > bókkoomm!*  
burn < 1PL > APPR  
‘If you don’t narrow down the fire properly (to the small spot in the center of the fire-place), (the fire) might light something, and we might burn.’ (Field notes 2006, DWD)

### 3.3.2. The morphosyntactic status of the support verbs

Y- ‘say’ and a’- ‘do’ are not limited to the ideophone-supporting function but also used on their own. For the independent use of y- ‘say’, see §4; for a’- ‘do’, see (10) (for the stem alternation a’- ~ at- ‘do’ see §3.3.3).

- (10) *M-á*                    *at-táyyoont?*  
what-M.ACC do-2SG.PROG  
‘What are you doing?’ (Geetaahun 2002: 178)

It is difficult to decide on the morphosyntactic status of ‘say’ and ‘do’, as ideophones and support verbs always immediately follow each other, no other constituents can intervene, and no morpheme can be suffixed to the ideophone or prefixed to the support verb.<sup>18</sup> I analyze y- ‘say’ and a’- ‘do’ as enclitic to the ideophones, and I mark the juncture with an equal sign. The support verbs are definitely not completely dependent on the ideophones (i.e. not suffixal), but they are not fully

<sup>18</sup> This is different from Sidaama, where, according to Kawachi (2007: 175), constituents can intervene between ideophones and support verbs.

independent words either. They retain their independent stress marking when inflected (§3.3.3) but tend to be reduced segmentally. The root consonant of *y-* ‘say’ is often hardly audible in non-careful speech. Regarding [ʔ]a’- ‘do’, the expected initial glottal stop, a phonetically determined boundary marker of vowel-initial words, is only realized in careful speech, e.g. *xóqq=a’-ú* ‘to raise’ is usually pronounced [t’ok’:a’ʔu] rather than [t’ok’:ʔa’ʔu] ‘raise’.<sup>19</sup> The ambiguous status of the support verb is also reflected in written Kambaata texts, where one of the most common sources of inconsistency is their representation as (in)dependent elements. Some authors generally write both support verbs as separate (e.g. Alemu 2016; Alamu & Alamaayyo 2017), others attach ‘do’ but are undecided about ‘say’ (e.g. Geetaahun 2002; Kambaata and Hadiyya Translation Project Hosaina 2005), still others write ‘say’ as separate but are undecided about ‘do’ (Desalegn 2020). All this indicates that native speakers are not sure about the support verbs’ status either; compare (11) and (12) from the same text.

- (11) (...) *ba’anch-áan* <*luqqitoot*> *lúqq=y-ítoot-i*  
 argument-F.LOC enter\_quickly.IDEO = say-2SG.NEG2-2SG.IMP  
 ‘(...) don’t enter quickly into an argument.’ (Kambaata Education Bureau 1989: 7.25)

- (12) (...) <*calba yitu’naan*> *calbá=y-itu’nnáan*  
 waver.IDEO = say-2SG.NEG4.CVB  
 ‘(...) without (you) wavering.’ (Kambaata Education Bureau 1989: 7.25)

The frozen middle suffix *-a’*, e.g. in *buqq(-)a’*- ‘uproot’ (Treis 2022: 23), may easily be mistaken as the *a’*- ‘do’ support verb and vice versa. Even though they are segmentally identical in non-careful speech, they differ in the morphophonological processes and the stress patterns they bring about (13)-(14) when inflected.

<sup>19</sup> Crass (2005: 228) mentions that ideophone and support verb often fuse into one phonological word in K’abeena. Since he does not analyze K’abeena stress in much detail, it is unclear whether there is a suprasegmental difference between regular verbs and ideophone-support constructions. The support verb ‘say’ is reduced and fused with the ideophone in Libido (liq), but it is consistently indicated in the transcription by Crass (n.d.: 30). For Alaaba, Schneider-Blum (2007: 277) reports about difficulties to tease regular verbs and ideophones apart.

(13) *tákk* = *a'*- ‘simply drop’ (IDEO)

3M.PRF = 3M.PFV: *tákk* = *a'-ée'u*

3F.PRF / 3F.PFV: *tákk* = *át-tee'u* / *tákk* = *at-tóo'u*

(14) *buqq(-)a'*- ‘uproot’ (VERB)

3M.PRF / 3M.PFV: *buqq(-)á'-ee'u* / *buqq(-)á'-o*

3F.PRF / 3F.PFV: *buqq(-)á'-ee'u* / *buqq(-)a'-óo'u*

### 3.3.3. The variants and the inflectional potential of the support verbs

The verbs *y-* ‘say’ and *a'*- ‘do’ are the most frequent but not the only possible support verbs for ideophones. In intransitive contexts, the verb *ih-* ‘become’ is sometimes attested where *y-* ‘say’ would be expected. It is unknown when *ih-* is preferred over *y-*, e.g. with *sá'mm* = *y-* ~ *sá'mm* = *ih-* ‘become quiet’, *xúnn* = *y-* ~ *xúnn* = *ih-* ‘smoke, fume’ or color ideophones, as in (15)-(16); when asked, consultants are fine with either support verb.<sup>20</sup>

(15) *Íchchi-yan*                      *hangaar-eemmá*                      *má'nn-it*  
 eat.3M.PFV.CVB-DS    scratch-3HON.PFV.REL    place-F.NOM  
*bíishsh* = *ík-k*<sup>21</sup>                                      *xid-ú*  
 red.IDEO = become-3F.PFV.CVB    hurt-M.ACC

(From a dictionary definition) ‘of a spot where it itched (and) that one scratched[:] to become red and hurt.’ (Alemu 2016: 126)

(16) *Qachch-ó*                      *áaz-u*                      *gambáll=y-án*                      *qax-ée*  
 beehive-F.GEN    interior-M.NOM    black.IDEO = say-3M.IPFV.CVB    extent-M.DAT  
*barg-í*                      *barg-í*                      *akkis-áamm*  
 add-1SG.PFV.CVB    add-1SG.PFV.CVB    smoke.CAUS1-1SG.IPFV

‘Until the interior of the beehive has become black, I continue smoking it.’  
 (AYZ2021-06-25\_006)

<sup>20</sup> See also the use of *ih-* ‘become’ with *xóphph* ‘jump’ in (17), an ideophone that usually combines with *y-* ‘say’ in my data. Speaker TH uses *ih-* ‘become’ more frequently than other speakers.

<sup>21</sup> *ík-k* is the result of a regular morphophonological assimilation process: *h + t > kk*.



A rare alternative intransitive support verb is *eeh-*; no independent use is attested or elicitable, but (historically) it might have been the base of the causative verb *ees-* ‘do’, a rare variant of *ass-* ‘do’ (Kambaata Education Bureau 1989: 125).

- (17) (...) *qo'rr-í*                      *xóqq = eekk-ó-o*<sup>22</sup>                      *maskoot-í*                      *al-íichch*  
           *gird-3M.PFV.CVB*    *rise.IDEO = ?-3M.PFV.REL-CIRC*    *window-M.GEN*    *top-M.ABL*  
*xóphph = íkk*                                      *had-áta*                      *láqq (...)*  
*jump.IDEO = become.3M.PFV.CVB*    *outside-F.ACC*    *head\_towards.3M.PFV.CVB*  
 ‘(...) when he had gotten ready and stood (up there), he jumped out from the window (...).’ (Frog Story 2007, TH)

In transitive contexts, the short verb *a'* ‘do’ is in free variation with the heavier verb *ass-* ‘do’; see (164) and (34). The verbs are entirely synonymous, but *a'* is much more frequent with ideophones and *ass-* much more frequent elsewhere.

- (18) *Kank-á <n>ka*                      *shínn~shínn = ass-í*  
           *this\_much-M.ACC <EMP>*    *RED~laugh\_heartily.IDEO = do-3M.PFV.CVB*  
*osal-siis-áyyoo-'nne = r-u*                      *m-áha-a-n?*  
*laugh-CAUS2-3M.PROG-2PL.OBJ = NMZ4-M.NOM*    *what-M.PRED-M.COP2-Q*  
 ‘What makes you (PL) laugh so heartily?’ (Kambaata Education Bureau 1989: 4.34)

The morphophonology and stress patterns of the support verb *a'* ‘do’ are unlike that of any other verb with a root-final glottal stop, as the comparison of the perfect and perfective paradigms in Table 6 show. If we focus on the shaded 3M and 3F forms, we see that the 3M forms of *a'* ‘do’ follow the model of *ga'* ‘call’. In contrast, the 3F forms of *a'* ‘do’ are unlike that of the two other verbs: Their stress patterns are similar to that of *ossa'* ‘lie down’, but we observe neither the *t*-deletion of the 3F markers *-tóo'u/'-tee'u* (3F.PFV/3F.PRF) that is characteristic of verbs like *ossa'* ‘lie down’, nor the insertion of an epenthetic vowel before the 3F markers *-tóo'u/'-tee'u* (3F.PFV/3F.PRF) that is characteristic of *ga'* ‘call’.

<sup>22</sup> The stem of the support verb *eeh-* is here realized as *eekk-*, because stem-final *h* undergoes gemination and fortition in certain forms of the perfect(ive) due to a regular morphophonological process.

	<b>a'- 'do'</b>		<b>ossa'- 'lie down'</b>		<b>ga'- 'call'</b>	
	PFV	PRF	PFV	PRF	PFV	PRF
1SG	<i>a'-éemm</i>	<i>a'-éemm</i>	<i>ossa'-óomm</i>	<i>ossá'-eemm</i>	<i>ga'-éemm</i>	<i>ga'-éemm</i>
2SG	<i>at-tóont</i>	<i>át-teent</i>	<i>ossa'-óont</i>	<i>ossá'-eent</i>	<i>ga'-itóont</i>	<i>ga'-íteent</i>
3M	<i>a'-ée'u</i>	<i>a'-ée'u</i>	<i>ossá'-o</i>	<i>ossá'-ee'u</i>	<i>ga'-ée'u</i>	<i>ga'-ée'u</i>
3F	<i>at-tóo'u</i>	<i>át-tee'u</i>	<i>ossa'-óo'u</i>	<i>ossá'-ee'u</i>	<i>ga'-itoo'u</i>	<i>ga'-ítee'u</i>
3HON	<i>a'-éemma</i>	<i>a'-éemmaa'u</i>	<i>ossa'-éemma</i>	<i>ossá'-éemmaa'u</i>	<i>ga'-éemma</i>	<i>ga'-éemmaa'u</i>
1PL	<i>a'-nnóomm</i>	<i>a'-nnéemm</i>	<i>ossa'-nnóomm</i>	<i>ossá'-nneemm</i>	<i>ga'-inóomm</i>	<i>ga'-íneemm</i>
2PL	<i>at-téenta</i>	<i>at-téentaa'u</i>	<i>ossa'-éenta</i>	<i>ossá'-éentaa'u</i>	<i>ga'-itéenta</i>	<i>ga'-itéentaa'u</i>

**Table 6:** Comparison of verbs with a root-final glottal stop (*ossa'- 'lie down'* is representative of all verbs with a frozen middle marker *-a'*).

The literature on Highland East Cushitic languages reports the use of ‘say’ and ‘do’ as support verbs in all languages (Table 7).

Language	Support verbs		Source
Kambaata	<i>y-</i> ‘say’ ~ <i>ih-</i> ‘become’	<i>a'-</i> ~ <i>ass-</i> ‘do’	(own data)
K'abeena	<i>y-</i> ‘say’	<i>a'-</i> ~ <i>ass-</i> ‘do’	(Crass 2005: 226)
Alaaba	( <i>y-</i> ) ‘say’	<i>a'-</i> ~ <i>ass-</i> ‘do’	(Schneider-Blum 2007: 277–278; 304–305)
Hadiyya	( <i>y-</i> ) ‘say’	<i>a'-</i> [ <i>&lt; *do'?</i> ]	(Hudson 1976: 273; Tadesse 2015: 346) <sup>23</sup>
Libido	( <i>y-</i> ) ‘say’	<i>aa'-</i> ‘receive’ [ <i>?</i> ] <sup>24</sup>	(Crass n.d.: 30)
Sidaama	<i>y-</i> ‘say’	<i>ass-</i> ‘do’	(Kawachi 2007: 174–179; Girum 2013: 44–45)
Gedeo <sup>25</sup>	<i>hiyy-</i> ‘say’	<i>ass-</i> ‘do’	(Eyob 2015: §4.1.3.1)
Burji <sup>26</sup>	<i>iy-</i> ‘say’	<i>iss-</i> ‘do’	(Tesfaye 2015: 198–200)

**Table 7:** Cognate support verbs for ideophones in Highland East Cushitic.

Whereas all Cushitic languages employ ‘say’ in intransitive contexts, the choice of the support verb in transitive contexts varies quite a bit. In Lowland East Cushitic,

<sup>23</sup> Tadesse (2015) does not provide any information on the support verbs of ideophones (or any other information on this word class). The data in his appendix shows, however, that potentially ideophonic transitive lexemes have a recurring *-a(a)'*-element, e.g. *witt'-* ‘gather’ (vi) – *witt'a'-* ‘collect’; see also the noun *witt'ite* ‘meeting’, which very likely contains the deideophonic nominalizer *-it* (§3.4.2). According to Hudson (1976: 273), the root consonant of ‘say’ is “omitted ordinarily” in Hadiyya, which might explain that is not written by Tadesse (cf. the situation in Libido and Alaaba). See also Hadiyya *huf-* ‘boil (vi)’ – *hufa'-* ‘boil (vt)’ and *hinc'-* ‘approach; become close (vi)’ – *hinc-a'-* ‘present (vt)’ in Hudson (1989: 282).

<sup>24</sup> I am not convinced that the support verb *aa'-* used in transitive contexts in Libido goes back to ‘accept, receive’, as claimed by Crass (n.d.: 30). Morphophonologically, it behaves like *a'-* in K'abeena and Alaaba, and it might (historically) originate from a ‘do’-verb in Libido, too.

<sup>25</sup> Gedeo: *drs*.

<sup>26</sup> Burjii: *bji*.

‘say’ pairs with ‘do’ in Oromo (orm) and Saho (ssy) (Amina 2013: 41–45; Banti & Vergari 2005), with ‘cause to say’ in Konso (kxc) and Gawwada (Ongaye 2013: 247; Tosco 2006: 888), with ‘give’ in Somali (Cabdulqaadir & Tosco 1998: 129) and with ‘put’ in Afar (Cohen et al. 2002: 228; Hassan Kamil 2015: 392–394). For the Central Cushitic language Xamtanga (xan), Darmon argues that the support verbs originate from ‘say’ and ‘cause to say’ (2015: 293–294). Similarly, in Beja (bej) (Vanhove 2017: 111–112, 147), ‘say’ and ‘cause to say’ accompany ideophones (or onomatopoeia). In the South Cushitic languages Iraqw (irk) and Alagwa (wbj), only ‘say’ is used with ideophones (Mous 1993: 228–229; 2016: 202–205).

Amharic (Ethiosemitic), the Ethiopian lingua franca and most important second language of Kambaata speakers, employs ‘say’ and ‘do’ as support verbs for ideophones (Meyer this volume).

### 3.3.4. The argument structure of ideophones

Most ideophones occur in a ‘say’/‘do’-pair. When native speakers are consulted, even ideophones that are only attested with ‘say’ in the data usually also permit – given an appropriate context – a combination with ‘do’ and vice versa. If both support verbs are permitted, ‘say’ is used in intransitive contexts, ‘do’ in transitive contexts. Semantically, one could characterize ideophones with ‘say’ as expressing non-causal events and ideophones with ‘do’ as expressing causal events. Paired ideophones allow the ‘do’-member to take an accusative object – see the accusative objects of ideophones in (7)-(9) – and to be passivized. The passive of ‘do’ is functionally equivalent to ‘say’ in ideophone-support constructions (19)-(20).

(19) *Kú*                      *xáh-u*                      *bóoc = y-ée = da*  
 A\_DEM1.M.NOM      thing-M.NOM      clarify.IDEO = say-3M.PFV.REL = COND  
*xúm-a-a*  
 good-M.PRED-M.COP2  
 ‘It’s good (/better) if this issue is clarified.’ (Field notes 2023, DWD)

(20) *Kú*                      *xáh-u*                      *bóoc = ass-amm-ó = da* (~ *bóoc = a'-amm-ó = da*)  
 A\_DEM1.M.NOM      thing-M.NOM      clarify.IDEO = do-PASS-3M.PFV.REL = COND  
*xúm-a-a*  
 good-M.PRED-M.COP2  
 ‘It’s good (/better) if this issue is clarified.’ (Field notes 2023, DWD)

One rarely comes across ideophones for which one of the support verbs is ruled out (unpaired ideophones). See, however, onomatopoeic ideophones for animal cries (Treis & Deginet 2024), e.g. *kutáakk=y-* ‘(of chicken) produce a sound which warns others of a predator’, which do generally not permit ‘do’. Support with ‘do’ was also ruled out for *hánda=y-* ‘be thankful, grateful, content’, *bíira=y-* ‘forgive’ and *quíuxo=y-* ‘do deliberately, willfully, on purpose, according to a plan’.<sup>27</sup> There is reason to assume that these three ideophones (and possibly others) are not intransitive, in spite of them combining with ‘say’. The evidence is most conclusive for *bíira=y-* ‘forgive’. In (167), the patient whose sin the agent forgives is the direct object of the ideophone; see the accusative marking of *nées* ‘us’. Through the passivization of ‘say’, the patient becomes the subject in (22).

- (21) *Magán-o, néés bíira=y-í-nne*  
 God-M.OBL 1PL.ACC forgive.IDEO = say-2SG.IMP-1PL.OBJ  
 ‘God, forgive us (our sins)!’ (Field notes 2022, DWD)

- (22) *Bíira=y-am-ámm<sup>28</sup> fanqáll tees-ú-u*  
 forgive.IDEO = say-PASS-PASS.3M.PFV.CVB return.3M.PFV.CVB now-M.OBL-ADD  
*ba’-íshsh-o*  
 go.bad-CAUS1-3M.PFV  
 ‘He was forgiven, (but then) he made a mistake again.’ (Field notes 2023, DWD)

Examples such as (167)-(22) show that the use of ‘say’ as a support verb is not sufficient proof that an ideophone is intransitive. While some ideophones lack a ‘do’-counterpart, the opposite configuration is also possible. In the semantic field of experiencer verbs, the use of ‘say’ is often not attested, sometimes explicitly excluded through negative evidence; see, e.g., *culú=a’-* ‘please (s.o.)’, *qúl~qúl=a’-* ‘make (s.o.) feel nostalgic, cause a longing feeling (in s.o.)’, *láf~láf=a’-* ‘make (s.o.) feel nauseous’, which take the stimulus as the subject and the experiencer as the direct object. Regarding their argument structure, ideophones thus seem to fall into (at least) four types:

<sup>27</sup> *Hánda* is not only an ideophone but also an interjection (§4.1), but the independent (interjectional) use of *bíira* and *quíuxo* was ruled out.

<sup>28</sup> Note that *y-* ‘say’ forms an irregular passive with *-am-am* (PASS-PASS), rather than simple *-am* (PASS).

- Type 1: Paired ideophone (most common): intransitive ideophone with ‘say’, transitive counterpart with ‘do’, e.g. *bóoc=y-* ‘be clarified’ – *bóoc=a’-* ‘clarify’; ‘do’-support passivizable.
- Type 2: Unpaired ideophones:
  - Type 2a: intransitive ideophone with ‘say’, no transitive counterpart with ‘do’, e.g. onomatopoeic ideophones for animal cries such as *búu=y-* ‘buzz (of bee)’ and bodily sounds such as *dú”u-dú”u=y-* ‘beat (of heart)’;
  - Type 2b: transitive ideophone with ‘say’, no counterpart with ‘do’, e.g. *búira=y-* ‘forgive’; ‘say’-support passivizable;
  - Type 2c: transitive ideophone with ‘do’, no counterpart with ‘say’, e.g. experiencer verbs such as *culú=a’-* ‘please (s.o.)’.

The support verbs ‘say’ and ‘do’ can undergo middle derivation with *-’/-aqq* and the reciprocal derivation with *-am/-aqq-am-* (Treis 2023b); see, e.g., the reciprocal forms *forgó~forgó=y-aqq-am-* ‘keep a distance from each other’, *gurgú=y-aqq-am-* ‘quarrel with each other’ and *háaf=y-aqq-am-* ‘forgive each other’ in Alamu (2022: 89, 152, 293; translation mine). The support verb *a’-* ‘do’ forms an irregular middle form *eecc-* ‘do for oneself’ (*\*a’-aqq-*), as in (168)-(24) from two different speakers.<sup>29</sup>

(23) *Éman*                      *xumm-ín*                      *gulub-í-kk*  
 congratulations.INTJ    peace-M.ICP                      knee-F.ACC-2SG.POSS  
*qu’mmí=eecc-ít*  
 bring\_together.IDEO = do.MID-2SG.PFV.CVB  
 (Addressed to new mother:) ‘Congratulations, you have brought your knees together again in peace!’ (AN2016-02-19\_001)

(24) *Ma’nn-íta*                      *xább=eecc-inim-bá’a*                      *y-éen (...)*  
 place-F.ACC    prepare.IDEO = do.MID-1PL.NIPFV-NEG1    say-3HON.PFV.CVB  
 ‘One says, “we have not (yet) prepared the place (for our benefit)”, (...)’  
 (EK2016-02-23\_003)

<sup>29</sup> The middle *eecc-* ‘do for oneself’ is based on a free variant of the ‘do’ verb, namely *ees-*, cf. paragraph before (17).

Finally, there remains the question of the syntactic relation between the ideophone and the support verb. In his analysis of Gawwada, Tosco considers the ideophones “direct objects” of the verb ‘say’ and ‘make say’ (2006: 890). For Kambaata, there is no evidence that ideophones are direct objects of their accompanying support verbs: they cannot be pronominalized or be passivized upon, and they are not accusative-marked – which are all obligatory features of a Kambaata direct object.

### 3.4. The morphology of ideophones

This section analyzes word formation processes that take ideophones as inputs and/or produce them as outputs. The first subsections deal with two semi-productive derivational processes for deideophonic resultative adjectives (§3.4.1) and action nouns (§3.4.2). I then move on to two processes that have both ideophonic inputs and outputs, namely reduplication (§3.4.3) and compounding (§3.4.4). The last sub-section (§3.4.5) discusses the rather marginal examples of ideophones that are derived through conversion from adjectives and pronouns.

#### 3.4.1. Deideophonic adjectivization

Resultative adjectives are derived from ideophones by *-eem*, e.g. *xóqq=y-* ‘rise’ > *xoqq-eem-á(ta)* ‘raised, elevated, at higher altitude’ and *hiríkk=y-* ‘be lowered, bend over, stoop’ > *hiríkk-eem-á(ta)* ‘lower(ed), at lower altitude’, as seen in (25).

- (25) *Lág-u*                      *xoqq-éem-ata*                      *ma’nn-éechch*  
 river-M.NOM      rise.IDEO-RES-F.OBL      place-F.ABL  
*hiríkk-eem-á = bb-a*    *zaaz-anó-o*  
 be\_lowered.IDEO-RES-M.ACC = PLC-M.ACC      flow-3M.IPFV.REL-NMZ1.M.ACC  
*ikk-ó = biiha*    (...)  
 become-3M.PFV.REL = REAS2  
 ‘Because rivers flow from a higher/elevated place towards a lower place (...)’  
 (Kambaata Education Bureau 1989: 8.71)

The resultative derivation is exclusive to ideophones, and no non-ideophonic inputs are attested. The resultative adjectives belong to A1, the largest adjectival declension;

consequently, their citation (= accusative) form ends in *-á* in the masculine and *-áta* in the feminine gender. Table 8 complements a list given in Treis (2008: 283–284).

Ideophone	English	Adjective	English
<i>qubúbb = y-</i>	‘become round, circular, ball-shaped’	<i>qububb-eem-á(ta)</i>	‘round, circular, ball-shaped’
<i>cagágg = y-</i>	‘become unwell, indisposed, a bit ill’	<i>cagagg-eem-á(ta)</i>	‘unwell, indisposed, a bit ill’
<i>síww = y-</i>	‘become thin (e.g. of body, voice)’	<i>siww-eem-á(ta)</i>	‘thin (e.g. of body, voice)’
<i>cerérr = y-</i>	‘trickle, flow in a trickle’	<i>cererr-eem-á(ta)</i>	‘trickling (e.g. of a jet of water); thin, elongated stretch (of land)’
<i>géxx = y-</i>	‘go straight, go directly, stand straight’	<i>gexx-eem-á(ta)</i>	‘straight, linear, direct’
<i>lácc = y-</i>	‘go slowly, slow down (vi)’	<i>lacc-eem-á(ta)</i>	‘slow (e.g. in walking, working)’
<i>salákk = y-</i>	‘become quiet, calm (of a place), attentive’	<i>salakk-eem-á(ta)</i>	‘calm, quiet, attentive’
<i>wikkíkk = y-</i>	‘cling (to s.th.), not be dissuaded (from s.th.)’	<i>wikkikk-eem-á(ta)</i>	‘stubborn, obdurate’
<i>qú’mm = y-</i>	‘be brought together/closer to each other; gather (vi)’	<i>qu’mm-eem-á(ta)</i>	‘narrow, constricted (e.g. opening of a pot)’
<i>(lúf~)lúf = y-</i>	‘become soft, malleable, fluffy’	<i>luf~luf-eem-á(ta)</i>	‘soft, malleable, fluffy (e.g. a mattress)’
<i>(qáf~)qáf = y-</i> -	‘become restless, hyperactive, irrepressible’	<i>qaf~qaf-eem-á(ta)</i>	‘restless, hyperactive, irrepressible’
<i>qól~qól = y-</i>	‘do too hastily, distractedly; meddle, interfere’	<i>qol~qol-eem-á(ta)</i>	‘distracted, too hasty; meddling, interfering’

**Table 8:** Deideophonic adjectives.

The resultative derivation takes simplex and reduplicated ideophones as input, as the last three lines of Table 7 and the second derived adjective in (26) illustrate.

- (26) *laal-ú-s*                      *qububb-eem-áan*                      [*fushsh-anó*]  
 fruit-F.ACC-3M.POSS      become.round.IDEO-RES-F.ICP      come\_out.CAUS1-3M.IPFV.REL  
*fatil-á*                      *wez-eennó*                      *luf~luf-eem-á*  
 thread-M.ACC      make-3HON.IPFV.REL      RED~become\_soft.IDEO-RES-M.ACC  
*hallaam-í*                      *algod-ahá-a*  
 soft\_interior\_of\_plant-M.GEN      plant-M.ACC-ADD  
*hallaam-u-sí-i*  
 soft\_interior\_of\_plant-M.ACC-3M.POSS-ADD  
 (Monolingual definition of *fuuttá* ‘(i.) cotton plant, (ii.) cotton’:) ‘[denotes] [i.] a plant that produces ball-shaped fruits/flowers [and] whose soft material is used for making threads and [ii.] the soft material itself.’ (Alemu 2016: 331 [corrected by DWD 2023])

### 3.4.2. Deideophonic nominalization

Action nouns are derived from ideophones by *-it(t)*, e.g. *xóqq=y-* ‘rise’ > *xoqq-itt-á* ‘height, elevation’, *sáww=y-* ‘think’ > *saww-itt-á* ‘thought, idea, (neolog.) sentence’; see also *xáphph=y-* ‘gather (vi)’ > *xaphph-itt-á* ‘gathering, assembly’, as in the neologism *xaphph-itt-í min-í* ‘synagogue (lit. house of assembly)’ of the Bible (Kambaata and Hadiyya Translation Project Hosaina 2005). If the ideophone ends in a short vowel, this vowel is dropped, e.g. *gambá=y-* ‘come across, encounter by chance/suddenly’ > *gamb-itt-á* ‘chance encounter’; after a long vowel, an epenthetic *h* is inserted, e.g. *múu=y-* ‘howl (of hyena)’ > *muu-h-itt-á* ‘time when hyenas howl at night’. Only one known *-it(t)*-derived noun is not based on an ideophone: *dang-itt-á* ‘unexpected event’ (28) is derived from one of the very few adverbs that the language has, namely *dángo* ‘unexpectedly’. All *-it(t)*-derived nouns belong to the masculine declension M1, ending in *-á* in their citation (= accusative) form.<sup>30</sup> The use of deideophonic action nouns is illustrated in (27)-(28).

<sup>30</sup> The two allomorphs *-itt* and *-it* are probably dialectal variants. In Alemu (2016), most deideophonic nouns have *-it*, while my consultants from the Duuraame area consistently use *-itt*. In Kambaata’s close relative K’abeena, the deideophonic nominalizer is *-it*, too; it is, however, marked



- (27) *Caam-á-i*                      *shobb-itt-á*                      *usúrr*                      *dagujj-óomm*  
 shoes-M.ACC-1SG.POSS    tie\_loosely-ACT-M.ACC    tie.1SG.PFV.CVB    run-1SG.PFV  
 ‘I tied my shoes loosely/quickly and ran off.’ (Field notes 2022, DWD)

- (28) *Tah-íchch-u*                      *dang-ítt-a*                      *waall-ó*                      *j-eechch-úta*  
 flies-SGV-M.NOM    unexpectedly-ACT-M.OBL    come-3M.PFV.REL    time-SGV-F.ACC  
*qeraa’rr-utá-a*                      *xámm = át-t*                      *af-fáa*  
 long-F.ACC-ADD    stick.IDEO = do-3F.PFV.CVB    seize-3F.IPFV.REL  
*arrab-i-sé-e*                      *mexx-é*                      *tokkós-s*  
 tongue-F.ACC-3F.POSS-ADD    single-MULT    shoot-3F.PFV.CVB  
*fush-sh-ít = ke’éechch*                      *xámm = át-t*  
 go\_out-CAUS1-3F.PFV.CVB = SEQ    stick.IDEO = do-3F.PFV.CVB  
*qúrc = at-táa’*,                      *tah-ichch-ú*  
 swallow.IDEO = do-3F.IPFV    flies-SGV-M.ACC  
 ‘When a fly comes unexpectedly, it (= the chameleon (F)) shoots out, at once, its long (tongue) and sticky, seizing tongue, makes (it) stick (to the tongue) and swallows (it), the fly.’ (TD2016-02-11\_001)

The examples in Table 9 complement a list presented in Treis (2008: 157–158).<sup>31</sup>

Ideophone	English	Noun	English
<i>báqq = y-</i>	‘wake up’	<i>baqq-itt-á</i>	‘dawn, daybreak’ <sup>32</sup>
<i>bóoc = a’-</i>	‘give a hint, give some indication, clarification’	<i>booc-itt-á</i>	‘hint, indication, clarification’
<i>cagágg = y-</i>	‘become unwell, indisposed, a bit ill’	<i>cagagg-itt-á</i>	‘being unwell, indisposition, being a bit ill’
<i>háff = y-</i>	‘pass, be perceived (of a faint smell)’	<i>haff-itt-á</i>	‘whiff, faint passing smell’

by *-i* in the citation form and thus grouped into a different declension than in Kambaata (Crass 2005: 84, 233).

<sup>31</sup> The *-it(t)*-derivation also takes reduplicated ideophonic inputs, as seen in *riph~riph-itt-á* (possibly): ‘up and down/wave movement’ in Kambaata Education Bureau (1989: 4.51). Unfortunately, the translation of the noun could not be verified and is, therefore, not included in the table.

<sup>32</sup> In Treis (2008: 157), the derived noun was translated as ‘waking up’; this translation was not confirmed and is here corrected.

Ideophone	English	Noun	English
<i>hazámm = a'</i>	'make (s.o.) doze off, make (s.o.) snooze briefly (esp. old people)'	<i>hazamm-itt-á</i>	'doze, snooze'
<i>hánda = y-</i>	'become thankful, grateful, content; become gratifying'	<i>hand-itt-á</i>	'thanks, gratitude, contentment'
<i>híqq = a'</i>	'cause to hiccup'	<i>hiqq-itt-á</i>	'hiccups'
<i>hó"á = a'</i>	'cause to vomit'	<i>ho"-itt-á</i>	'vomiting'
<i>múgg = a'</i>	'lower; cause to have diarrhea'	<i>mugg-itt-á</i>	'diarrhea'
<i>qú'mm = y-</i>	'be brought together/closer to each other; gather (vi)'	<i>qu'mm-itt-á</i>	'collection; meeting, assembly, session'
<i>shóbb = y-</i>	'be loosely tied'	<i>shobb-itt-á</i>	'loose tie'
<i>silímm = a'</i>	'cause to take a nap, cause to have a short/interrupted sleep'	<i>silimm-itt-á</i>	'nap, short/interrupted sleep'
<i>xéff~xéff = y-</i>	'have small spots, dots'	<i>xeff-itt-á</i>	'spot, dot'
<i>yák~yák = a'</i>	'do haphazardly, do too quickly and carelessly, struggle to do'	<i>yak~yak-itt-á</i>	'haphazard, too quick and careless manner'
<i>zammamm = a'</i>	'cause to doze off, cause to be half asleep'	<i>zammamm-itt-á</i>	'doze, half-sleep'

Table 9: Deideophonic nouns.

Apart from the two semi-productive deideophonic processes discussed above, there are a small number of nouns and verbs which are otherwise derived from ideophones. Among them are nouns derived by *-eenn*, e.g. *bizz-eenn-á* 'up(wards), North' < *bízz = y-* 'rise'; *mugg-eenn-á* 'down(wards), South' (used in (6)) < *múgg = y-* 'descend (vi)'; *cererr-eenn-á* 'trickle, rivulet, small rest of a drink' < *cerérr = y-* 'trickle'; *ilill-eenn-á* 'ululation' < *ilíli = y-* 'ululate' and *tilill-eenn-á* 'telephone' < *tililí = y-* 'ring'. However, unlike the *-it(t)*-derivation, the non-productive *-eenn*-derivation is not restricted to ideophonic inputs (Treis 2008: 165).

Finally, Treis & Deginet (2024: ex. 11) report a handful of verbs derived from onomatopoeic ideophones, e.g. *kaakk-ees-* 'cackle (of hen which is about to lay an egg)', based on the ideophone *káakk = y-* with the same meaning (see two similar examples from K'abeena in Crass 2005: 148). Most inputs of the semi-productive *-ees*-derivation in Kambaata are nouns (Treis 2022: 24).





synonymous to the reduplicated form. A similar judgment was, for instance, made with respect to (*haráxx~*)*haráxx*=*y*- ‘stride, walk with quick, long steps’ in (32), where the reduplicated form was considered to be replaceable by a simplex form (here the entry in Alemu 2016: 450 is the simplex form).

- (32) *Haráxx~haráxx*=*y-itán*                      *már-i*,                      *hawwarr-ókkoo-he*  
 RED~stride.IDEO = say-2SG.IPFV.CVB    go-2SG.IMP    become\_dark-3M.APPR-2SG.OBJ  
 ‘Walk fast(er) or it might become dark on us!’ (Field notes 2022, DWD)

#	Ideophone	English
1	<i>bárk~bárk</i> = <i>y</i> -*	‘flap (of wings)’
2	<i>béxx~béxx</i> = <i>y</i> -	‘show up, start to come out (of plural subject)’
3	<i>birxí~birxí</i> = <i>y</i> -	‘writhe, twist, wriggle about (e.g. to free oneself)’
4	<i>bolóng~bolóng</i> = <i>a</i> '-♦	‘move one’s eyes (up/down, left/right)’
5	<i>chílċ~chílċ</i> = <i>y</i> -	‘clink, jingle (of keys, coins, small bells on a horse’s neck)’
6	<i>cóphph~cóphph</i> = <i>y</i> -	‘drip (e.g. of leaky water-tap)’
7	<i>dúbb~dúbb</i> = <i>a</i> '-	‘pound (coffee leaves)’
8	<i>farcá~farcá</i> = <i>a</i> '-♦	‘crumble, break up (sticky product, e.g. fermented tobacco, enset food)’
9	<i>forgó~forgó</i> = <i>y</i> -	‘move a bit away from each other, keep a distance from each other, leave some space (e.g. so as not to bump into one another) (of plural subject)’
10	<i>gíphph~gíphph</i> = <i>y</i>	‘stop briefly, pause (when walking or talking) (of plural subject)’
11	<i>haráxx~haráxx</i> = <i>y</i> -	‘stride, walk with quick, long steps’
12	<i>hurbú~hurbú</i> = <i>a</i> '-*	‘make (the belly) grumble’
13	<i>láf~láf</i> = <i>a</i> '-	‘make (s.o.) feel nauseous’
14	<i>óorin~óorin</i> = <i>y</i> -*	‘encourage a child to take a laxative; abet, goad, prompt (s.o.) repeatedly’
15	<i>qúl~qúl</i> = <i>a</i> '-	‘make (s.o.) feel nostalgic, cause a longing feeling (in s.o.) (e.g. of person or home that is missed)’
16	<i>qulcú~qulcú</i> = <i>y</i> -♦	‘slosh back and forth (of water in a container when carried)’
17	<i>qumbú~qumbú</i> = <i>a</i> '-*	‘drum, make (a drum) sound’
18	<i>shínn~shínn</i> = <i>y</i> -	‘laugh heartily (of plural subject/continuously)’
19	<i>xéff~xéff</i> = <i>y</i> -*	‘have/be covered by small spots, dots’
20	<i>yák~yák</i> = <i>a</i> '-	‘do haphazardly, do too quickly and carelessly, struggle to do’

**Table 11:** Reduplicated ideophones

(\*only attested as reduplicated in my data and in Alemu 2016; ♦only attested as reduplicated in my data, not attested in Alemu 2016; rest: attested as simplex and as reduplicated).

In most cases, however, the reduplication of ideophones seems to be triggered by the iterative nature of the event that they express. Iterativity can be inherent in the event, as in #1 ‘flap’, #14 ‘abet, goad, prompt repeatedly’ and #17 ‘drum’ in Table 11, which are (probably) always reduplicated. For other ideophones, an iterative realization of the event is very common; see *cóphph=y-* expressing a single dripping event (one drop) vs. the reduplicated *cóphph~cóphph=y-* (#6 in Table 11) expressing a repeated dripping event (multiple drops), e.g. of a leaky water-tap.

A slightly different meaning, a subtype of event plurality, is reflected in (33), where the reduplicated ideophone expresses repeated (possibly cautious) attempts to do something. According to a consultant, the reduplicated ideophone in (33) is (near) synonymous with the iterative/attenuative-derived ideophone in (34).

- (33) (...) *iséta da'll-ít xóqq~xóqq = at-táa*  
 3F.ACC do\_soon-3F.PFV.CVB RED~rise = do-3F.IPFV.REL  
*mateemínn-it yoo-ba'í = dda (...)*  
 shortage\_of\_manpower-F.NOM exist.3-NEG1.REL = COND  
 ‘(...) if there is no shortage of manpower (in the household) that makes her (= the new mother) soon try to rise up (again from childbed, in order to work) (...).’ (EK2016-02-23\_002)

- (34) *Mann-í bác-u huj-ée xóqq = ass-ans-áno-se*  
 people-M.GEN shortage-M.NOM work-3F.DAT rise.IDEO = do-ITER-3M.IPFV-3F.OBJ  
 ‘The shortage of man(power) makes her gradually rise up for work (again).’  
 (Field notes 2023, DWD)

Apart from event plurality (iterativity), reduplication also serves to express participant plurality;<sup>33</sup> see #2 ‘show up’, #9 ‘keep a distance’, #10 ‘pause’ and #18 ‘laugh heartily’ in Table 11, whose reduplicated forms are attested with plural subjects. The use of the simplex and reduplicated forms of ideophone #9 is illustrated in (35)-(36).

<sup>33</sup> In the literature on related languages, ideophone reduplication is said to express repeated events, events expressed by plural subjects and prolonged (continuous) events (see Amina 2013: 36–37 on Oromo; Crass 2005: 292–293 on K’abeena; Girum 2013: 70 on Sidaama; Eyob 2015: §9.3 on Gedeo).

- (35) *Qah-ú* < *n* > *ka*                      *zakk-ú*                      ***forgó*** = *y-ít*  
 small-M.ACC < EMP >    behind-M.ACC    move\_away.IDEO = say-2SG.PFV.CVB  
*úurr-i*  
 stand-2SG.IMP  
 (Possible context: Command addressed to a person who has crossed the yellow line) ‘Move back a little bit and stand (there)!’ (Field notes 2022, DWD)

- (36) *Gomb-aqq-an-ténoochch-iyye*,                      ***forgó*** ~ ***forgó*** = *y-itéen*  
 bump-MID-PASS-2PL.NEG2-2PL.IMP    RED ~ move\_away.IDEO = say-2PL.PFV.CVB  
*uurr-iyyé*  
 stand-2PL.IMP  
 (Possible context: command addressed to a crowd) ‘Please, do not bump into each other, keep your distance/stand a bit away from each other!’ (Field notes 2022, DWD)

#### 3.4.4. Compound ideophones

In my fieldwork data and in written sources, there is a small number of bipartite ideophones which result from compounding. Overall, compounding is not a very productive word formation process in Kambaata, the most frequent compound type are numeral-plus-noun compounds (Treis 2017: 349–351), whereas, for instance, noun-noun compounds like *min-ann-á* (house-father-M.ACC) ‘head of the household’ and noun-verb compounds like *oz-gal-á* (dinner-spend\_the\_night-M.ACC) ‘something to eat for dinner’ are rare. Even more exceptional are verb-verb compounds; the only attested verbally inflecting example, *birs-awwa’nn-(aqq-)am-* (precede.CAUS1-follow.MID-(MID-)PASS) ‘follow one another’, is possibly a recent calque from Amharic.

The use of a compound ideophone is shown in (37); see (40) for another example.

- (37) *Tá*                      *wotar-ch-úta*                      *aass-íha*                      *tí*                      *hárr-it*  
 A\_DEM1.F.ACC    foal-SGV-F.ACC    give-M.DAT    A\_DEM1.F.NOM    donkeys-F.NOM  
***shambá.qambá*** = *y-itáni-yan*    *wotár-ch-u-s*  
 bustle\_around\_in\_joyful\_expectation-3F.IPFV.CVB-DS    foal-SGV-F.NOM-DEF  
*káar-t (...)*    *fírúx-x*    *ir-úchch*  
 become\_suspicious-3F.PFV.CVB    gallop-3F.PFV.CVB    land-M.ABL  
*ba’-óó’u*  
 disappear-3F.PFV

‘When the donkeys were bustling around in joyful expectation to hand the foal over (to the hyenas), the foal became suspicious, (...) galloped away and disappeared from the face of the earth.’ (Kambaata Education Bureau 1989: 4.34)

Table 12 lists some compound ideophones from Alemu (2016) that are confirmed by at least one other source.<sup>34</sup>

#	Ideophone	English	Origin
1	<i>bízz-múgg = a’-</i>	‘make have the stomach flu; make vomit and make have diarrhea’	<i>bízz = a’</i> ‘raise; make vomit’ + <i>múgg = a’</i> ‘lower; cause to have diarrhea’
2	<i>fógg-xóqq = y-</i>	‘be busy (especially of a host who needs to prepare and serve food, entertain the guests etc.)’	<i>fógg = y-</i> ‘bend over/down’ + <i>xóqq = y-</i> ‘rise’
3	<i>kimbíll-biciríqq = y-</i>	‘tumble, turn over [?]’	<i>kimbíll = y-</i> / <i>biciríqq = y-</i> ‘return back, swerve suddenly’
4	<i>míkk-míll = y-</i>	‘budge, move a bit (lit. and fig.)’ - especially used in negative contexts: not move even a bit; see (6)	<i>míkk = y-</i> (meaning synonymous to compound); independent use and meaning of 2 <sup>nd</sup> part not confirmed
5	<i>qashá-qombó = y-</i>	‘snap, respond impatiently, irritably, brusquely (and thus socially inappropriately)’	<i>qashá = y-</i> (meaning synonymous to compound); independent use and meaning of 2 <sup>nd</sup> part not confirmed
6	<i>shambá-qambá = y-</i> -	‘bustle around, be busy preparing (in joyful expectation)’	independent use and meaning of parts not confirmed
7	<i>shóott-gáff = y-</i>	‘be emotionally moved, agitated, fidget (because of anger, worries)’	<i>shóott = y-</i> ‘get up abruptly’ + <i>gáff = y-</i> ‘sit down for a little while’

Table 12: Compound ideophones.

Based on the meanings of their parts, Desalegn (2020: 68) proposes to categorize Kambaata compound ideophones into combinations of synonyms (e.g. #3 in Table

<sup>34</sup> Ideophone compounding is also reported for K’abeena (Crass 2005: 233) and Sidaama (Girum 2013: 46).



12), combinations of antonyms (e.g. #1 in Table 12) and combinations of an ideophone with a semantically empty counterpart (e.g. #5 in Table 12).<sup>35</sup>

Alternatively, the compounds could be classified according to whether their meaning is predictable. Compound #1 in Table 12 has a transparent origin and expresses the meanings of its parts. In contrast, the examples in #2 and #7 show that the meaning is not necessarily compositional: the concept of busyness (#2) is rendered as a combination of (repeated) bending over and rising up again; the concept of being agitated (#7) is expressed as a combination of (repeated) getting up abruptly and sitting down briefly. As far as semantic idiosyncrasies are concerned, ideophonic compounds are not different from other lexicalized compounds in the language.

### 3.4.5. Ideophone formation by conversion

Most Kambaata property concepts (apart from colors) can be expressed by adjectives and inchoative-stative verbs of the same root, e.g. *danq-á(ta)* ‘deaf’ – *danq-* ‘become deaf’, *itis-á(ta)* ‘expensive’ – *itis-* ‘become expensive’, *qug-á(ta)* ‘raw’ – *qug-* ‘become raw’. There is no conclusive evidence making it possible to decide whether adjectives are derived (through conversion) from inchoative-stative verbs or vice versa, or whether property lexemes should be assumed to be undetermined for word class.<sup>36</sup> A similar conversion relation can be observed between color adjectives and ideophones (and not verbs). Color ideophones come in intransitive/transitive pairs, e.g. *gamball-á(ta)* ‘black’ and *gambáll=y-* ‘become black’ / *gambáll=a’-* ‘blacken’, *wojj-ú(ta)* ‘white’ and *wójj=y-* ‘become white’ / *wójj=a’-* ‘whiten’; for the use of color ideophones in context see (15)-(16). Here again, we can ask whether color adjectives are derived from ideophones or vice versa. This time, there seems to be an argument in favor of the direction being adjective > ideophone. Unlike other ideophones, color ideophones cannot be the input of the adjectivizing derivation by *-em* (§3.4.1) or of the nominalizing derivation by *-it(t)* (§3.4.2). Instead, color nouns are derived from the adjective by the productive deadjectival nominalizer *-im*, see *gamball-á(ta)* ‘black’ – *gamball-im-áta* ‘blackness’, just like *qug-á(ta)* ‘raw’ – *qug-*

<sup>35</sup> In K’abeena, Crass (2005: 293–294) considers examples such as #4 and #6 in Table 12 to be “echo word” formations, in which the second part of the compound does not have a meaning on its own but simply “echoes” the first part phonologically, i.e. produces a (near) copy.

<sup>36</sup> Recall that adjectives are cited in their accusative case form, i.e. the endings *-á(ta)*, *-ú(ta)* etc. of cited adjectives are inflectional, not derivational morphemes.

*im-áta* ‘rawness’, which leads me to assume that the color adjective is the base and the ideophone derived.

We also find two ideophones that share a root with a pronoun. The manner demonstratives *hitt-íta* (F) ‘like this’ and *hatt-íta* (F) ‘how’ (Treis 2019) have two corresponding ideophonic demonstrative pairs, *hítt=y-* ‘do like this’ / *hítt=a-* ‘do (s.th.) like this’ and *hátt=y-* ‘become how, turn out where’ / *hátt=a-* ‘do how, do what (to s.th.)’. The transitive members of the pairs are seen in (38)-(39); see (53) for *hítt=ih-*.

- (38) (...) *xit-á*                      *ang-áan*                      *hítt=a'-éen*  
 soot-M.ACC    hand-F.ICP    do\_like\_this.IDEO = do-3HON.IPFV.CVB  
*usur-éen=ké'* (...)  
 tie-3HON.IPFV.CVB = SEQ  
 ‘(...), like this [GESTURE], one ties the (bunch of grass with the) soot with the hand and then (...).’ (EK2016-02-23\_002)

- (39) *Xon-é*                                      *buud-á-s*                                      *hátt=at-tóo-la?*  
 aforementioned-F.GEN    horn-M.ACC-DEF    do\_how.IDEO = do-3F.PFV-MIT  
 ‘What did she do with (i.e. where has she put) that horn?’ (TD2016-02-11\_001)

Finally, one ideophone, *lamúbb=y-* ‘fall into two parts’ (40), is based on the stem of the denumeral noun *lam-ú=bb-a* (two-M.ACC=PLC-M.ACC) ‘into two parts’, a combination of the numeral *lam-ú* ‘two’ (M.ACC) and the place nominalizer =*b(b)* (Treis 2008: 243–247). Here it is undisputable that the ideophone is derived (through conversion) from the denumeral and not vice versa.

- (40) (...) *hiliq-óon*                      *háam-u-s*  
 shock-M.LOC    chest-F.NOM-3M.POSS  
*fūiqq-táaqq=y-ít*  
 be\_ripped\_apart.IDEO-be\_ripped\_apart.IDEO = say-3F.PFV.CVB  
*boq-u-sí-i*                                      *áaz-u*                                      *qixx-ú*  
 head-F.NOM-3M.POSS-ADD    interior-M.NOM    equal-M.ACC  
*lamúbb=y-ít*<sup>37</sup>                                      *farshá=y-itóo-s*  
 fall\_into\_two\_parts = say-3F.PFV.CVB    break\_apart.IDEO = say-3F.PFV-3M.OBJ

<sup>37</sup> The ideophonic form is here phonetically realized as *lamúbb=ít*, with the *y* of the ‘say’-verb omitted, but the independent stress on the support verb is retained.

(Description of a person's severe state of shock) '(...) his chest was ripped apart in shock and his head, the inside, broke neatly into two parts.' (Frog Story 2007, TH)

If compared to Amharic (Meyer this volume), where ideophones can be very productively derived from verbs, the derivation of ideophones is a marginal phenomenon in Kambaata and limited to some cases of conversion.

### 3.5. The semantics of ideophones

With regard to their semantics, ideophones range from the semantically broad (Table 13) to the semantically very specific and the evocation of complex scenes, e.g. *hokkóbb=y-* '(e.g. of a goat) stand on the hind legs and lean the front legs against (e.g. a tree)'.

Intransitive	English	Transitive	English
<i>biddíqq=y-</i>	'become flat, even, level; spread (vi) out'	<i>biddíqq=a'-</i>	'make flat, even, level; spread (vt) out'
<i>búrr=y-</i>	'fly'	<i>búrr=a'</i>	'make fly'
<i>cál~cál=y-</i>	'hang (vi) freely'	<i>cál~cál=a'-</i>	'hang (vt) freely'
<i>dákk=y-</i>	'be hidden, concealed'	<i>dákk=a'-</i>	'hide, conceal'
<i>fógg=y-</i>	'bend over/down'	<i>fógg=a'-</i>	'make bend over/down'
<i>húnc=y-</i>	'come close(r)'	<i>húnc=a'-</i>	'bring close(r)'
<i>lácc=y-</i>	'go slow, slow down (vi)'	<i>lácc=a'-</i>	'do (s.th.) slowly, slow (s.th.) down'
<i>qúrc=y-</i>	'be swallowed'	<i>qúrc=a'-</i>	'swallow'
<i>sá'mm=y-</i>	'become quiet, fall silent'	<i>sá'mm=a'-</i>	'silence (s.o.)'
<i>shígǵ=y-</i>	'become shocking, disgusting, revolting'	<i>shígǵ=a'-</i>	'shock, disgust, revolt (s.o.)'
<i>xább=y-</i>	'do (vi) well, in a straight manner'	<i>xább=a'-</i>	'do (s.th.) well, in a straight manner, prepare (s.th.)'
<i>xíshsh=y-</i>	'be tightened, strengthened, deepened'	<i>xíshsh=a'-</i>	'do (s.th.) tightly, strongly, deeply'

**Table 13:** Illustrative list of semantically fairly broad ideophones.

Verbal concepts could be said to be expressed arbitrarily by verbal or by ideophonic lexemes in Kambaata; compare, e.g., *búrr=y-* 'fly' (IDEO) vs. *dagud-* 'run' (VERB); *cál~cál=y-* 'hang (vi) freely' (IDEO) vs. *olaal-* 'hang (vi) freely; lean (e.g. of a tree





Alemu (2016) that are paraphrased as being a type of *usur-* ‘tie’, selected examples that I was able to verify are given in Table 14.

Verb	English
<i>gaajj-</i>	‘tie up the back legs (of a cow for milking, so that it does not kick out)’
<i>kuttubb-</i>	‘tie up two legs (esp. of equines and small livestock so that they do not wander far)’
<i>laangoob-</i>	‘tie up to the right and left (of a cow in its pen so that it cannot turn around)’ (synonym: <i>laambees-</i> )
<i>saankaal-</i>	‘(Alemu:) tie the neck to the front leg / (my field notes:) tie a front to a back leg (of an equine so that it does not wander far)’
<i>humbuub-</i>	‘tie up, wrap (esp. butter in enset leaf sheaths)’
<i>qurunnees-</i>	‘tie up in a bundle (so that it can be carried on the back, esp. enset food, cabbage, wood)’
<i>qirqaab-</i>	‘tie up a load (esp. a load on a donkey)’
<i>wiriir-</i>	‘tie a cloth around the opening of a pot’
<i>farshimm-</i>	‘tie up in a sheaf’
<i>wixtix-</i>	‘tie tightly together (e.g. two poles)’

Table 14: Verbs from the semantic field of tying.

At present (and until further study), there is little reason to assume that Kambaata ideophones are semantically any different from verbs.

### 3.6. Ideophone frequency and size of the ideophone word class

Ideophones are not restricted to a particular text type but found in narrative texts, procedural texts, conversations, proverbs (43) and written texts (e.g. schoolbooks).

- (43) *Gassim-á* [kórx-o] *luucc-ee-sé* *guumm-íchch-ut*  
morning-M.ACC jump-M.OBL fail.MID-3M.PFV-3F.OBJ.REL *duikers-SGV-F.NOM*  
*tú'mm = y-itán* *hos-sáa'u*  
try\_in\_vain.IDEO = say-3F.IPFV.CVB spend\_the\_day-3F.IPFV  
(Proverb) ‘A duiker who failed when jumping in the morning tries the whole day in vain.’ (Alamu & Alamaayyo 2017: 56 [correction by DWD])

One even finds ideophones in biblical texts (Table 15). The Gospel of John (Kambaata and Hadiyya Translation Project Hosaina 2005) contains 13 ideophone types and 73 tokens (in 11,851 words), i.e. 1 ideophone in ~162 words. If compared to the number of ideophones in a conversation (1,135 words, 1 ideophone in ~31 words) in Table 16, one can formulate the working hypothesis that natural conversational data has a higher ideophone ratio than translated data.

Ideophone type	English	Tokens
<i>biddíq[q] = y-</i>	‘be spread out, be stretched out, be	1
<i>biddíq[q] = a’-</i>	levelled’	1
	‘spread out, stretch out, level’	
<i>bínn = y-</i>	‘be dispersed, be scattered’	3
<i>bínn = a’-</i>	‘disperse, scatter’	1
<i>búll = ih-</i>	‘become greyish/ashen’	1
<i>hátt = ih-</i>	‘become how, turn out where’	9
<i>hátt = a’-</i>	‘do how, do what (to s.th.)’	3
<i>hínc = y-</i>	‘come close(r)’	13
<i>hínc = a’-</i>	‘bring close(r)’	2
<i>hiríkk = y-</i>	‘be lowered, bend over, stoop’	9
<i>hítt = a’</i>	‘do (s.th.) like this’	1
<i>qúrc = a’-</i>	swallow (s.th.)	2
<i>tú = y-</i>	‘spit’	1
<i>úff = a’-</i>	‘blow, breathe (into s.th.)’	1
<i>wíll = y-</i>	‘sneak away, disappear (from s.o.’s view)’	1
<i>xáphph = y-</i>	‘gather (vi)’	5
<i>xáphph = a’-</i>	‘gather (vt)’	5
<i>xóqq = y / xóqq~xóqq = y-</i>	‘rise’	5/2
<i>xóqq = a’-</i>	‘raise’	7
<b>13</b>		<b>73</b>

Table 15: Ideophone types attested in the Gospel of John (11,851 words).

Ideophone type	English	Tokens
<i>bárr = y- / búrr y-</i>	‘fly’	2
<i>bolóng~bolóng = a’-</i>	‘move one’s eyes (up/down, left/right)’	1
<i>címm = y-</i>	‘slump down, shrink (vi)’	3
<i>hátt = ih-</i>	‘become how, turn out where’	2
<i>hátt = a’-</i>	‘do how, do what (to s.th.)’	1

Ideophone type	English	Tokens
<i>hítt = ih-</i>	‘do (vi) like this’	5
<i>kuchúchch = y-</i>	‘contract, shrink, coil up’	2
<i>láaxx = a’- / láaxx-láaxx = a’-</i>	‘do (s.th.) quickly’	1/1
<i>míkk.míll = y-</i>	‘budge, move a bit’	2
<i>qúrc = a’-</i>	‘swallow (s.th.)’	4
<i>sáww = y-</i>	‘think’	5
<i>xább = y-</i>	‘do (vi) well, in a straight manner’	1
<i>xámm = a’-</i>	‘make (s.th.) stick’	4
<i>xíshsh = y-</i>	‘be tightened, strengthened, deepened’	1
<i>xóphph = y-</i>	‘jump’	1
<i>xóqq = y-</i>	‘rise’	1
<b>15</b>		<b>37</b>

**Table 16:** Ideophone types attested in a conversation about a picture book (TD2016-02-11\_001; 12:10 min; 1,135 words; prompt: Carle 1984).

A varied repertoire of ideophones is also a manifestation of a person’s narrative skills. One of my consultants (TH) with a strong literary interest wrote up a Frog Story in Kambaata (based on Mayer 1969), a text with which he wanted to demonstrate his story-telling talent. The resulting text of 1,187 words (*Fe’llóncho, hakkánne yóont?* ‘Frog, where are you?’, unpublished) contains no less than 46 ideophone types and 64 tokens (i.e. 1 ideophone in ~19 words). The ideophone frequency in the (crafted) Frog Story is thus about twice as high as in the conversational data displayed in Table 16. A quantitative study of ideophones across text types/genres and across individual speakers would make an interesting subject for future research.

In order to approach the question of how large the ideophone word class is, Alemu’s comprehensive (2016) dictionary was taken as an empirical base. The author marks all ideophonic roots that obligatorily combine with either ‘say’ or ‘do’ by an asterisk (\*) (2016: xxii).<sup>40</sup> A search of all \*-marked roots yielded 853 entries, of which most are paired ideophones.<sup>41</sup> Quantifying the number of verb roots in the

<sup>40</sup> Interestingly, he does not invent a word class term for \*-marked roots. In contrast, neologisms are created for all other Kambaata word classes (e.g. *shoosaaww-á* ‘verb’).

<sup>41</sup> Alemu’s data contained originally 968 asterisks, but needed to be cleaned a bit (which I failed to do for Treis 2022). 51 lexemes which were said to combine with ‘say’ and ‘do’ but where the asterisk was forgotten were added to the count. 166 asterisks that occurred in accompanying texts outside the actual word list were subtracted. Note that this picture is still skewed, as the monolingual definitions



dictionary would require considerable effort, because not only roots but also all derived stems, any combination with a support verb and even verbs in cross-references are marked as “(sh)” for *shoosawwá* ‘verb’. To be able to provide a rough estimate of the number of verbal roots in the dictionary, I counted them on 20 randomly chosen pages (pp. 368-389). The result was an average of 2.05 verb roots per page, which led to an estimate of 2388 verb roots for the whole dictionary (1,165 pages). This means that, among all lexemes expressing verbal concepts, ideophones amount to an estimated 26% (853/3241) and verbs to about 74% (2388/3241).

Borrowing into the ideophone word class is possible and observed in my field work data.<sup>42</sup> As Amharic ideophones (see Meyer this volume) are also invariant and morphosyntactically integrated with ‘say’ and ‘do’, Kambaata speakers can simply borrow the ideophonic root and inflect them with the corresponding Kambaata support verb, as (44) attests. In (44), speaker ED first uses the Amharic loan ‘sit down’ < ቀላጭ ኣለ *q<sup>w</sup>əčč alä* (Kane 1990: vol. 1: 843) with the Kambaata support verb, then translates the ideophone into Kambaata.

- (44) (...) *tées-u*                    ***qúcc = y-éen***  
           now-M.OBL    sit\_down.IDEO[AMH] = say-3HON.PFV.CVB  
*afuu'll-éen*                    *kánk-u*                    *has-is-áno*  
 sit\_down 3HON.PFV.CVB    this\_much-M.NOM    want-CAUS1-3M.IPFV  
 ‘(...) one sits down, sits down (and says:) “This much is needed.”’ (EK2016-02-23\_003)

#### 4. Other ‘say’-constructions

This section gives an introduction into the expression of reported speech in general and of reported interjections in particular. It points out the fuzzy boundary between reported interjections and ideophone-support constructions. In the second part,

---

contain a number of ideophones which do not yet occur as entries (and are hence not \*-marked). No dictionary can ever be complete, and a number of ideophones attested in my fieldwork data are absent from Alemu (2016), e.g. *bolóng~bolóng = a'* ‘move one’s eyes (up/down, left/right)’ and *kuchúchch = y-* ‘contract, shrink, coil up’, to name but two examples from Table 16.

<sup>42</sup> See also the discussion on ideophone borrowing in Crass (2005: 234) on K’abeena.

constructions with ‘say’ for the expression of transitory or vestigial tastes and smells are portrayed. Here the coverbs of ‘say’ are not ideophonic but nominal.<sup>43</sup>

#### 4.1. Reported speech

The verb *y-* ‘say’ is used in the matrix clause of reported speech constructions. It either follows the reported speech as the only utterance verb, see (45), or functions as a coverb to a semantically narrower, superordinate utterance verb, e.g. *kul-* ‘tell’, *fanqashsh-* ‘reply’, *xa’mm-* ‘ask’ (46).

- (45) (...) *mát-o*            *yaa’-áan*            *mánn-u*            *afuu’ll-ée*  
                   one-M.OBL    meeting-M.LOC    people-M.NOM    sit\_down-3M.PFV.REL  
*j-áata*            [*hannó*            *á’nnu*            *baad-ú*            *maassa’-éen*  
 time-F.ACC    please.INTJ    2HON.NOM    land-M.ACC    bless-2PL.PFV.CVB  
*jammarr-é*] REPORTED SPEECH    *y-éeni-yan*            *nubáach-ch-u*  
 start-2PL.IMP                    say-3HON.PFV.CVB-DS    elders-SGV-M.NOM  
*baad-ú*            *maassa’-íi*            *jammarr-áno*  
 land-M.ACC    bless-M.DAT    start-3M.IPFV

‘(...) when people are sitting (together) in a meeting, one says, “Please, start to bless the land!”, and an elder starts to bless the land.’ (AN2016-02-19\_001)

- (46) (...) *híkka*            *man-ch-ú*            [*mát-it*            *gambáll-at*  
                   A\_DEM2.M.ACC    person-SGV-M.ACC    one-F.NOM    black-F.NOM  
*hóol-ch-ut*            *kannín*            *hig-góo’?*            *xuud-déentaa’*  
 sheep-SGV-F.NOM    P\_DEM1.M.ICP    pass-3F.PFV    see-2PL.PRF  
*íkke?*] REPORTED SPEECH    *y-í*                    *xa’mm-áno*  
 PAST                    say-3M.PFV.CVB    ask-3M.IPFV

‘(...) he (who looks for his animal) asks that man (who passes him): “Has a black ewe passed by here? Have you seen (it)?”’ (AN2016-02-19\_002)

The underived verb *y-* ‘say’ has three arguments, (i) the speaker, expressed minimally by a subject index, (ii) the addressee, optionally encoded by a

<sup>43</sup> Due to reasons of space, other constructions in which (a more or less bleached) ‘say’ is used are left out of the discussion, e.g. the purpose, benefactive focus, internal impulse and ‘at least’-constructions.

pronominal object suffix on ‘say’ and/or by an independent direct object noun phrase (see the accusative-marked *híkka manchú* ‘that man’ in (46)), and (iii) the speech report. If the reported speech is pronominalized (e.g. ‘She didn’t say that’), the pronoun is also accusative-marked, which makes *y-* ‘say’ a ditransitive verb.

The boundary between reported speech and ideophone-support constructions is fuzzy in two areas. Firstly, the small subclass of onomatopoeic ideophones (Treis & Deginet 2024), which contains, among others, animal cries like *humbáa=y-* ‘moo (of cattle)’ in (47) and bodily sounds like *buhhú=y-* ‘cough, produce a coughing sound’, might still be interpretable as a reported animal cry and a reported bodily sound.

- (47) *Kú*                      *sá’-u*                      *humbáa=y-ú*                      *batíshsh-ee’u*  
 A\_DEM1.M.NOM    cow-M.NOM              moo.IDEO = say-M.ACC    do\_much-3M.PRF  
 ‘This cow has mood a lot/too much.’ (Field notes 2022, DWD)

Secondly, the fine line between quoted interjections and ideophones is difficult to draw. Members of both word classes are morphologically invariant, but interjections, unlike ideophones, constitute utterances of their own. Thus one could simply consider all invariant coverbs in ‘say’-constructions, e.g. *shiinchaallé* ‘never again!’ in (48), to be interjections if they are also attested as independent utterances, which is the case for *shiinchaallé* ‘never again!’ in (49).

- (48) *Shiinchaallé*    *y-í*                      *gíbb-ee’u*  
 never\_again.INTJ    say-3M.PFV.CVB    refuse-3M.PRF  
 ‘He said “Never again!” and refused (to do it).’ (Field notes 2023, DWD)

- (49) *Shiinchaallé,*    *lan-kíi*                      *híkka=b-á*                      *mar-áam-ba’a*  
 never\_again.INTJ    two-ORD.DAT    A\_DEM2.M.ACC = PLC-M.ACC    go-1SG.IPFV-NEG1  
 ‘Never again, I won’t go there a second time!’ (Field notes 2023, DWD)

However, this approach is not entirely satisfactory in cases where the meaning of a construction consisting of an interjection plus ‘say’ is not the sum of its parts, or, said differently, where the combination has become conventionalized, has extended its meaning and is in most contexts no longer interpretable as a direct speech report. An example which illustrates this point is the interjection/ideophone *hánda* ‘(I/we/someone/it are/is well,) thanks to God’. The only possible addressee of the

speaker's thanks is God; people are not thanked with *hánda*. In exchanges about one's well-being (in the broadest sense, i.e. physical, psychological, economical), *hánda* constitutes an utterance of its own and is thus best interpreted as an interjection (50).

(50) [A:] *Wées-it iill-ítee-ndo?*

enset-F.NOM reach-3F.PRF-Q

[B:] *Hánda (isí) <sup>44</sup> iill-ítee'u*

thanks\_to\_God.INTJ3 M.DAT reach-3F.PRF

[A:] 'Are the enset (plants) ready (to be harvested)?' – [B:] 'Thanks to God, they are ready.' (Field notes 2023, DWD)

The combination of *hánda* and *y-* 'say' is in principle interpretable as 'say "Thanks to God"', but in the vast majority of cases it is best translated as 'be thankful, be grateful; be content; be gratifying'. Speakers do not interpret the examples in (51)-(52) as direct quotes of the subject, which is especially evident when the subject is an animal, as in (52). One might thus prefer not analyze *hánda* as an interjection in these contexts but as an ideophone that takes 'say' as a support verb.

(51) *Hánda = y-í gáll-ee-haa*

be\_grateful.IDEO = say-3M.PFV.CVB pass\_the\_night-3M.PRF.REL-M.COP2

(Speaking about someone's attitude/character) 'He is (someone) who is content (with his life, with what he has)/he is a positive person (lit. he is (someone) who passes the night gratefully).' (Field notes 2023, DWD)

(52) *Habank-á ichchí-i hánda = y-áano-ba'a*

how\_much-M.ACC eat.3M.PFV.CVB-ADD be\_grateful.IDEO = say-3M.IPFV-NEG1

(Speaking about a cow with a bad feed conversion ratio) 'No matter how much it eats, it is not gratifying.' (Field notes 2022, DWD)

A similar point can be made for the pain and grief interjection *áayye* 'ouch!, I am so sad/sorry (about the loss)!'. In combination with 'say', *áayye = y-*, it translates as 'call out in pain; call out in sadness (at a funeral, in the home of a deceased)'. Likewise, the interjection *ekkú* 'okay' commonly combines with 'say' to express *ekkú = y-* 'agree, obey', irrespective of whether the agreement has actually been

<sup>44</sup> *Hánda* can govern a dative constituent; the optional *isí* 'for him' refers to God.



Table 17 gives an overview of all taste and smell constructions attested in my fieldwork data with different speakers; some of the examples were overheard, others elicited.<sup>47</sup> All of them combine with *y-* ‘say’; combinations with *a-* ‘do’ are unattested and were ruled out by a consulted native speaker.

Ideophone	English	Origin
<i>babbáro~babbáro=y-</i>	‘have a taste of pepper’	<i>babbar-úta</i> (F) ‘hot Ethiopian spice mixture’
<i>gotí ále~ále=y-</i>	‘smell of hyena’s body (e.g. a person who does not wash)’	<i>got-á</i> (M) ‘hyena’, <i>al-í</i> (M) ‘body’
<i>háqqa~háqqa=y-</i>	‘have a bland taste, be boring (of a talk)’	<i>haqq-á</i> (M) ‘wood’
<i>haráqe~haráqe=y-</i>	‘smell of brandy (e.g. a person who is drunk)’	<i>haraq-íta</i> (F) ‘brandy’
<i>hixíchcho~hixíchcho=y-</i>	‘have a taste of lemongrass (e.g. cheese)’	<i>hixichch-ú</i> (M) ‘lemongrass’
<i>hóolla~hóolla=y-</i>	‘smell of sheep (e.g. a person who has been in the sheep pen)’	<i>hooll-áta</i> (F) ‘sheep’
<i>íchcha~íchcha=y-</i>	‘smell of food (e.g. a room)’	<i>ichch-áta</i> (F) ‘food’
<i>má’lle~má’lle=y- / má’lla~má’lla=y-</i>	‘have a bland taste, lack spices’	<i>ma’ll-á</i> (M) / <i>ma’ll-íta</i> (F) ‘bland food’
<i>maxíne~maxíne=y-</i>	‘have a taste of salt’	<i>maxin-íta</i> (F) ‘salt’
<i>shá’lla~shá’lla=y-</i>	‘smell of cow dung (e.g. a person who has mucked out the cowshed)’	<i>sha’ll-á</i> (M) ‘cow dung’
<i>wilíile~wilíile=y-</i>	‘smell of smoke (e.g. clothes of a person who has sat beside the fire)’	<i>wilíil-íta</i> (F) ‘smoke’
<i>wó’a~wó’a=y-</i>	‘taste watery (e.g. bad potatoes, bland oranges)’	<i>wo’-á</i> (M) ‘water’
<i>wóxe~wóxe=y-</i>	‘smell of sauce (e.g. the kitchen)’	<i>wox-íta</i> (F) ‘sauce’
<i>xágo~xágo=y-</i>	‘smell of spices (e.g. a person who has seasoned butter)’	<i>xag-íta</i> (F) ‘spices’

**Table 17:** Transitory and vestigial tastes and smells.

The construction might represent a metaphorical extension of the reported speech construction, in which a participant exuding a taste/smell is expressed as a speaker calling out the source of a taste/smell (note that vocatives are oblique case-marked).

<sup>47</sup> Of all examples presented in Table 17, only *má’lle~má’lle=y-* translated as ‘taste naïf’ and *maxíne~maxíne=y-* translated as ‘become brackish’ are found as entries in Alemu (2016: 691, 706).

## 5. Ideophones – word class of its own or subclass of verbs?

The present paper has taken a language-internal morphosyntactic definition of ideophones as a point of departure and then studied the form, the meaning and the use of ideophones across recorded, written and elicited data. Even though there are no productive mechanisms to generate ideophones (ideophone formation through compounding and conversion is very limited), it has become clear that ideophones are an open word class with a potentially infinite number of members, as attested by the ~850 ideophonic lexemes provided in Alemu (2016) and the adhoc integration of ideophones from Amharic into Kambaata utterances (§3.6). It now remains to be discussed whether it is more appropriate to consider ideophones as a separate word class or to subsume them under verbs as a subtype. As we will see, there are arguments for both positions.

Ideophones are distinct from verbs in that they obligatorily need a support verb to be morphosyntactically integrated into an utterance; verbs do not combine with ‘say’ and ‘do’ to form complex predicates. Ideophones have lexically determined stress, while verbal stems are undetermined for stress – stress marking is introduced through inflectional morphology. Idiolectal phonological variation is much more common for ideophones than for verbs. Kambaata has two derivational processes, the adjectivizing, resultative derivation and the action noun derivation that take ideophones as input, while verbal inputs are unattested. Ideophone reduplication is much more prominent than verb reduplication (or more appropriate: repetition).

In contrast, ideophones and verbs also share a number of features. Most importantly, ideophone-support constructions and verbs are used in the same syntactic functions as predicates in main and subordinate clauses. Ideophones usually come in intransitive/transitive pairs (‘say’-‘do’-pairs) and their argument structure maps onto that of intransitive and transitive verbs – notwithstanding that ‘say’ in its independent use as a speech verb is not intransitive but transitive. If we leave aside the small number of onomatopoeic ideophones, there is little to no evidence that ideophones are phonologically or phonotactically unusual. Regarding their semantics, §3.5 has cast doubt on the initial assumption that ideophones are semantically narrower than (or otherwise different from) verbs.

Based on the commonalities of ideophones and verbs, one could justifiably postulate a word class of verbs, which divides into two formal subtypes, true (non-periphrastic) verbs and periphrastic verbs. True verbs can be inflected directly,

periphrastic verbs require a semantically empty support verb. A division into two subtypes of verbs might even be supported by evidence from language-internal variation and change as well as comparative evidence – recall that the Kambaata support verbs are enclitic in nature and that the segmental substance of *y-* ‘say’ is about to be lost. At present, the stress patterns (and the unique morphophonology of the ‘do’-verb) are often the most reliable indicator for identifying ideophone-support constructions (i.e. periphrastic verbs). Also in closely related Highland East Cushitic languages, the consonant root of the ‘say’ verb is on its way to being lost. One could speculate that the lack of attention paid to ideophones in the literature on Hadiyya is attributable to the fact that the ‘say’ verb is already completely absorbed and that the ‘do’-verb is interpreted as a transitivity derivational morpheme. Overall, the support verbs are losing their independent status.

Having justified the existence of ideophones as a word class or sub-word class on language-internal ground, it remains to be determined whether Kambaata ideophones fit the comparative concept proposed by Dingemanse (2019: 16), namely “an open lexical class of marked words that depict sensory imagery”. As the openness of the word class and their structural markedness have been proven, I restrict my comments to the question of whether Kambaata ideophones “depict” and have “sensory meanings”. Only a subset of Kambaata ideophones seems iconic (see Treis & Deginet 2024 on sound-imitative ideophones and §3.4.3 on the iconic functions of reduplication); usually the mappings between forms and meaning are arbitrary. Nevertheless, as Dingemanse (2019: 19) points out for other languages, also Kambaata native speakers are adamant that the form of an ideophone is suggestive of its meaning. Dingemanse attributes this to the depictive nature of ideophones, which “invites and affords the construal of iconic mappings between form and meaning” (Dingemanse 2019: 18). Kambaata ideophones can express sounds (e.g. *xixí* ‘crackle when being roasted on the fire (of small seeds)’), movement (e.g. *túk=y-* ‘walk behind each other in large groups’), visual patterns (e.g. *xéff~xéff=y-* ‘have/be covered by small spots’), other sensory perceptions like texture (e.g. *líf~líf=y-* ‘become soft, malleable, fluffy’) and color (e.g. *búll=y-* ‘become greyish, ashen’), inner feelings and cognitive states (e.g. *cagágg=y-* ‘become unwell, indisposed, a bit ill’; *kulúl=a’-* ‘make feel dizzy’, *qúl~qúl=a’-* ‘make (s.o.) feel nostalgic’) – i.e. Kambaata covers the whole implicational hierarchy of semantic fields proposed by Dingemanse (2012: 663) – but I find it difficult to judge whether



Kambaata ideophones have closer ties to sensory perception than verbal lexemes. A more fine-grained semantic analysis of ideophones and verbs is definitely required.

Of course, this is not the only gap that remains. This paper hopefully paves the way for in-depth investigations of some of the aspects presented above. I was, for instance, neither able to do justice to the (possibly) substantial Kambaata-internal regional variation in the form and meaning of ideophones nor to the (presumably) significant differences in the frequency of use of ideophones across text types and across speakers. The degree of cognacy of ideophones across closely related languages is also a question that would merit to be investigated in the future.

### Acknowledgements

Work on this article was supported by a public grant overseen by the French National Research Agency (ANR) as part of the program “*Investissement d’Avenir*” (ANR-10-LABX-0083). It contributes to the IdEx Université de Paris (ANR-18-IDEX-0001). I am indebted to my language assistant Deginet Wotango Doyiso (abbreviated DWD in the examples above) for the intensive and fruitful discussions of the analyses proposed here, and to Alemu Bantu Atara (also spelt: Alamu Banta Ataara) for granting me access to his voluminous dictionary. I am grateful to all the Kambaata speakers I have been working with in the past years and who have contributed data that was analyzed for this paper; special thanks go, in alphabetical order, to Abel Yonas Zekarias (AYZ), Aman Nuriye (AN), Belachew Nuramo (BN), Ermias Kenore (EK), Teshome Dagne (TD), Tessema Handiso (TH) and Yoseph Yonas Zecharias (YYZ). Thanks to Martine Vanhove, the anonymous reviewers and the co-editors of this special issue for helpful comments on an earlier version of this paper.

### Abbreviations

1 = first person	nominalization	CAUS2 = long causative
2 = second person	ADD = additive	CIRC = circumstantial
3 = third person	ADJ = adjective	COND = conditional
A <sub>1</sub> = adjectival	AMH = Amharic	COP2 = ascriptive, identificational copula
ABL = ablative	APPR = apprehensive	-(h)a(a) (M) / -ta(a) (F)
ACC = accusative	BDV = benedictive	CVB = converb
ACT = action	CAUS1 = short causative	

DAT = dative	MIT = mitigator	PL = plural
DEF = definite	MULT = multiplicative	PLC = place nominalization
DEM1 = proximal demonstrative	NEG1 = standard negator	POSS = possessive
DEM2 = medial demonstrative	NEG2 = imperative negator	PRED = predicative
DS = different subject	NEG4 = converb negator	PRF = perfect
EMP = emphasis	NEG5 = relative negator	PROG = progressive
F = feminine	NIPFV = non-imperfective	Q = question marker
GEN = genitive	NMZ1 = nominalization with -V	REAS2 = reason clause marker = <i>biiha</i>
HON = honorific; impersonal	NMZ4 = nominalizer = <i>r-</i>	RED = reduplication
ICP = instrumental-comitative-perlative	NOM = nominative	REL = relative
IDEO = ideophone	OBJ = object pronoun	RES = resultative derivation
IMP = imperative	OBL = oblique	SEQ = sequential
ITER = iterative	P_ = pronominal	SGV = singulative
INTJ = interjection	PART = discourse particle	SIM = similitive
MID = middle	PASS = passive	vi/vt = intransitive / transitive verb
	PAST = past, counterfactuality	VOC = vocative
	PFV = perfective	

## References

- Alamu Banta & Alamaayyo G/Tsiyoon. 2017. *Hambarrichcho yaanata: Kambaatissa-Amaarsa hayyo'ooma yannaakkata* [Spices of Mount Hambarrichcho: Kambaata-Amharic proverbs]. Addis Ababa: Addis Ababa University.
- Alamu Banta Ataara. 2022. *Daambiliqa: Kambaatissa finka* [Intellectual advancement: Kambaata metaphors]. Duuraame: [n. p.].
- Alemu Banta Atara [Alamu Banta Ataara]. 2016. *Kookaata: Kambaatissa-Amaarsa-Ingiliizissa laaga doonnuta* [Kambaata-Amharic-English dictionary]. Addis Ababa: Berhanena Selam Printing.
- Amina Zaro. 2013. *Ideophones in Oromo*. Master Thesis, University of Addis Ababa.
- Appleyard, David L. 2001. The verb “to say” as a verb ‘recycling device’ in Ethiopian languages. In Andrzej Zaborski (ed.), *New data and new methods in Afroasiatic linguistics. Robert Hetzron in memoriam*, 1–11. Wiesbaden: Harrassowitz.
- Banti, Giorgio & Moreno Vergari. 2005. A sketch of Saho grammar. *Journal of Eritrean Studies* 4(1–2). 101–131.
- Cabdulqaadir Salaad Dhoorre & Mauro Tosco. 1998. 111 Somali ideophones. *Journal of African Cultural Studies* 11(2). 125–156.

- Carle, Eric. 1984. *The mixed-up chameleon*. New York: Harper Collins.
- Central Statistical Agency. 2007. Ethiopia: Population and housing census of 2007. <http://catalog.ihnsn.org/index.php/catalog/3583> (Last access: 4 January 2020).
- Cohen, David, Marie-Claude Simeone-Senelle & Martine Vanhove. 2002. The grammaticalization of “say” and “do”: An areal phenomenon in East Africa. In Tom Güldemann & Manfred von Roncador (eds.), *Reported discourse. A meeting ground for different linguistic domains*, 227–251. Amsterdam: John Benjamins.
- Crass, Joachim. 2005. *Das K’abeena. Deskriptive Grammatik einer hochlandost-kuschitischen Sprache*. Cologne: Rüdiger Köppe.
- Crass, Joachim. n.d. [Collection of transcribed, glossed and translated Libido texts].
- Darmon, Chloé. 2015. *A morphosyntactic description of Xamtanga, an Agaw (Central Cushitic) language of the northern Ethiopian highlands*. Doctoral dissertation, Université Lumière Lyon 2.
- Desalegn Haile Arficho. 2020. *Ideophones in Kambaata*. Doctoral dissertation, Addis Ababa University.
- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6(10). 654–672.
- Dingemanse, Mark. 2019. “Ideophone” as a comparative concept. In Kimi Akita & Prashant Pardeshi (eds.), *Ideophones, mimetics and expressives*, 13–56. Amsterdam: John Benjamins.
- Eyob Kelemework Wolde. 2015. *A grammar of Gedeo: A Cushitic language of Ethiopia*. Doctoral dissertation, Osmania University, Hyderabad.
- Ferguson, Charles A. 1976. The Ethiopian language area. In M. Lionel Bender, J. D. Bowen, R. L. Cooper & C. A. Ferguson (eds.), *Language in Ethiopia*, 63–76. London: Oxford University Press.
- Geetaahun Heellebo Baachchoore. 2002. *Kambaatissa afgale maxaafa* [Collection of Kambaata oral literature]. Tungu: Kambaatina Xambaaro Zoonaan, Dooyyoganni Woradaan, Maare-Ammaachcho Waato.
- Girum Tesfaye. 2013. *Ideophones in Sidaama: Documentation and description*. Master Thesis, University of Addis Ababa.
- Güldemann, Tom. 2005. Complex predicates based on generic auxiliaries as an areal feature in Northeast Africa. In F. K. Erhard Voeltz (ed.), *Studies in African linguistic typology*, 131–154. Amsterdam: John Benjamins.
- Hassan Kamil, Mohamed. 2015. *L’afar: Description grammaticale d’une langue couchitique (Djibouti, Érythrée et Éthiopie)*. Doctoral dissertation, Inalco, Paris.

- Hudson, Grover. 1976. Highland East Cushitic. In M. Lionel Bender (ed.), *The non-Semitic languages of Ethiopia*, 232–77. East Lansing, MI: African Studies Center, Michigan State University.
- Hudson, Grover. 1989. *Highland East Cushitic dictionary*. Hamburg: Buske.
- Kambaata and Hadiyya Translation Project Hosaina. 2005. *Qarichcho Yesuus Kiristoositannee Yohannis xaaffo Mishiraachchi Maxaafa / Latin version of the Gospel of John in Kambaata language*. Addis Ababa: The Bible Society of Ethiopia.
- Kambaata Education Bureau. 1989. *Kambaatissata: Rosaanchi maxaafa* [Kambaata language: School book]. 8 vols. Hawassa: Southern Nations, Nationalities, and Peoples Regional State.
- Kane, Thomas Leiper. 1990. *Amharic-English dictionary*. 2 vols. Wiesbaden: Harrassowitz.
- Kawachi, Kazuhiro. 2007. *A grammar of Sidaama (Sidamo), a Cushitic language of Ethiopia*. Buffalo: State University of New York at Buffalo (Doctoral Dissertation).
- Leslau, Wolf. 1945. The influence of Cushitic on the Semitic languages of Ethiopia: A problem of substratum. *Word* 1(1). 59–82.
- Mayer, Mercer. 1969. *Frog, where are you?* New York: Dial Press.
- Mous, Maarten. 1993. *A grammar of Iraqw*. Hamburg: Buske.
- Mous, Maarten. 2012. Cushitic. In Zygmunt Frajzyngier & Erin Shay (eds.), *Afroasiatic languages*, 342–422. Cambridge: Cambridge University Press.
- Mous, Maarten. 2016. *Alagwa - a South Cushitic language of Tanzania: Grammar, texts and lexicon* (Cushitic and Omotic Studies 7). Cologne: Rüdiger Köppe.
- Ongaye Oda Orkaydo. 2013. *A grammar of Konso*. Utrecht: LOT.
- Schneider-Blum, Gertrud. 2007. *A grammar of Alaaba. A Highland East Cushitic language of Ethiopia*. Cologne: Rüdiger Köppe.
- Tadesse Sibamo Garkebo. 2015. *Documentation and description of Hadiyya (a Highland East Cushitic language of Ethiopia)*. Master Thesis, University of Addis Ababa.
- Tesfaye Baye Assefa. 2015. *A descriptive grammar of Burji*. Doctoral dissertation, University of Addis Ababa.
- Tosco, Mauro. 2006. The ideophones in Gawwada. In Siegbert Uhlig (ed.), *Proceedings of the XVth International Conference of Ethiopian Studies, Hamburg, July 20-25, 2003*, 885–892. Wiesbaden: Harrassowitz.
- Treis, Yvonne. 2008. *A grammar of Kambaata. Part 1: Phonology, morphology, and non-verbal predication*. Cologne: Rüdiger Köppe.

- Treis, Yvonne. 2012. Kambaata children's songs and word games: Or what can we learn about the grammar of Kambaata by analysing marginal literature genres? *Afrikanistik Online*. 2012.
- Treis, Yvonne. 2017. "They are only two, like the teats of a donkey": Kambataa numerals revisited. In Raija Kramer & Roland Kießling (eds.), *Mechthildian approaches to Afrikanistik: Advances in language based research on Africa: Festschrift für Mechthild Reh*, 339–366. Cologne: Rüdiger Köppe.
- Treis, Yvonne. 2019. Simulative and equative demonstratives in Kambaata. *Faits de Langues* 51(1). 75–201.
- Treis, Yvonne. 2022. Verbal derivation in Kambaata (Cushitic), with a focus on the encoding of noncausal/causal alternations. *Linguistique et langues africaines* 8(2) (Special issue "The noncausal-causal alternation in African languages", ed. by Sebastian Dom, Leora Bar-el, Ponsiano Sawaka Kanijo & Malin Petzell). 1–36. <https://doi.org/10.4000/lla.4535>.
- Treis, Yvonne. 2023a. [13.] Kambaata. In Ronny Meyer, Bedilu Wakjira & Zelealem Leyew (eds.), *The Oxford handbook of Ethiopian languages*, 198–226. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198728542.013.13>.
- Treis, Yvonne. 2023b. Reflexive constructions in Kambaata. In Katarzyna Janic, Nicoletta Puddu & Martin Haspelmath (eds.), *Reflexive constructions in the world's languages*, 155–184. Berlin: Language Science Press. <https://dx.doi.org/10.5281/zenodo.7861660>.
- Treis, Yvonne & Deginet Wotango Doyiso. 2024. Onomatopoeias in Kambaata. In Lívia Körtvélyessy (ed.), *Onomatopoeias* (Comparative Handbooks of Linguistics), 81–92. Berlin: De Gruyter Mouton.
- Vanhove, Martine. 2017. *Le bedja*. Leuven: Peeters.
- Zaborski, Andrzej. 1991. Ethiopian language subareas. In Stanisław Piłaszewicz & Eugeniusz Rzewuski (eds.), *Unwritten testimonies of the African past. Proceedings of the International Symposium held in Ojrzeń n. Warsaw on 07-08 November 1989*, 123–134. Warsaw: Wydawnictwa Uniwersytetu Warszawskiego.

## CONTACT

yvonne.treis@cncs.fr

# Ideophones and verbal constructions with the verb ‘say’ in Amharic

RONNY MEYER

INALCO & CNRS, UMR 8135 LANGAGE, LANGUES ET CULTURES D’AFRIQUE

Submitted: 16/02/2023 Revised version: 7/02/2024

Accepted: 8/02/2024 Published: 12/06/2024



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

## Abstract

Amharic (Ethiosemitic) has verbal predicates of two types: (i) verbs derived from a consonantal root and (ii) complex predicates consisting of an invariant coverb followed by an inflecting light verb, most commonly the verb *alä* ‘say’. Although complex predicates have been discussed in several descriptions of Amharic, little attention has been paid to the different word classes that occur as coverbs and the differences in their morphosyntactic behavior. This is also true for ideophones, which are often mentioned in the discussion of complex predicates, but without stating whether they form a distinct lexical class in Amharic and how this class can be characterized. This article describes the various elements that function as coverbs in complex predicates and provides a language-internal morphosyntactic definition of ideophones as a word class in Amharic. It is argued that only an ideophonic coverb can form noncausal/causal complex predicate pairs with the light verbs *alä* ‘say’ and *adärrägä* ‘do’. In addition to basic ideophonic lexemes, Amharic can also derive ideophones from verbal roots through dedicated templates expressing intensification or attenuation of the event denoted by the root. Amharic ideophones rarely occur on their own, but are typically morphosyntactically integrated by light verbs. Finally, a comparison of different types of coverbs in Amharic with those in other Ethiosemitic languages shows that the languages that are geographically close to or in close contact with Amharic have a similarly large variety of coverbs, while the Ethiosemitic languages that are further away have a more limited inventory.

**Keywords:** Amharic; Ethiosemitic; complex predicates; coverb; verb ‘say’; ideophone

## 1. Introduction

Amharic (amh; Afroasiatic, Semitic)<sup>1</sup> is an Ethiosemitic language and a lingua franca of Ethiopia, a country with an estimated population of 110 million inhabitants in 2018. Ethiopia has over 80 languages belonging to two traditional language phyla: Afroasiatic which includes various Cushitic, Semitic and Omotic languages, and Nilo-Saharan (see Eberhard et al. 2022: 7). Approximately 57 million people speak Amharic in Ethiopia, with 25 million being second language speakers (Eberhard et al. 2022: 13). The language is native to the Amhara Regional State in northern Ethiopia, as well as to the Ethiopian capital Addis Ababa and many other cities throughout the country. Due to its extensive use in primary education and as the working language of the Ethiopian government, Amharic has spread beyond its native-speaking region (see Meyer 2011a). The language is relatively well described, with its first grammar and dictionary dating back to 1698 (Ludolf 1698a; 1698b). It has a rich literary tradition, particularly since the late 19<sup>th</sup> century (see Meyer 2011b: 1178–1180).

Amharic and over ten other languages constitute the Ethiosemitic branch within the Semitic language family, which belongs to the Afroasiatic language phylum (see Huehnergard & Pat-El 2019: 3–4; Hudson 2013: 6–36). Hetzron (1972: 119; 1977: 15–24) classifies the Ethiosemitic languages into three major groups, as shown in Figure 1 with minor modifications (for critical points regarding this classification, see Meyer 2018: 94–95).

### North Ethiosemitic

†Geez (gez), Tigre (tig), Tigrinya (tir), Dahaalik (dlk)

### Transversal Ethiosemitic

- a. Amharic (amh), Argobba (agj)
- b. Harari (har),  
East Gurage [EG]: Zay (zwa), Wolane (wle), Silt’e (stv)

### South Ethiosemitic

- a. †Gafat (gft)

---

<sup>1</sup> The languages cited are classified genealogically according to Glottolog and marked with their ISO-code 639-3 upon first mention.

- b. Gunnän Gurage [GG]
  - i. Northern Gurage: Muher (sgw), Kistane (gru), Dobbi (gru)
  - ii. Mesqan (mvz)
  - iii. Chaha group: Chaha, Gumer, Ezha, Gura (all sgw)
  - iv. Inor group: Inor, Enär, Gyeta, Endegagn (all ior),  
†Mesmes (mys)

**Figure 1:** Major branches of Ethiosemitic (Afroasiatic, Semitic)

The geographical distribution of the Ethiosemitic languages is shown in Figure 2.

As for the general typological profile, the common constituent order is S(ubject)–O(bject)–V(erb), head nouns always follow their modifiers, but case relators may precede, follow, or enclose their complement. Amharic is a nominative–accusative language in which only the accusative is marked obligatorily on definite nouns (1b), whereas indefinite nouns as primary object are most typically unmarked (1a). Verbs obligatorily index their subject and may optionally also index the primary object (1b), or an applied object (1c).<sup>2</sup> Differential object marking is common.

- (1) a. *almaz mäkina fät'-ättf*  
 Almaz.F car sell\PFV-SBJ.3SF  
 'Almaz sold a car.'
- b. *almaz mäkina-w-n fät'-ättf(-iw)*  
 Almaz.F car-DEF.M/POSS.3SM sell\PFV-SBJ.3SF(-OBJ.3SM)  
 'Almaz sold the/his car.'
- c. *tilant mäkina-w-n fät'-ättf-ibb-ijñ*  
 yesterday car-DEF.M/POSS.3SM sell\PFV-SBJ.3SF(-MIL-OBJ.3SM)  
 'Yesterday, she sold the/his car to my detriment.'

---

<sup>2</sup> The Amharic data is transcribed and transliterated with IPA symbols, except for the open-mid central vowel, which is represented by *ä*.



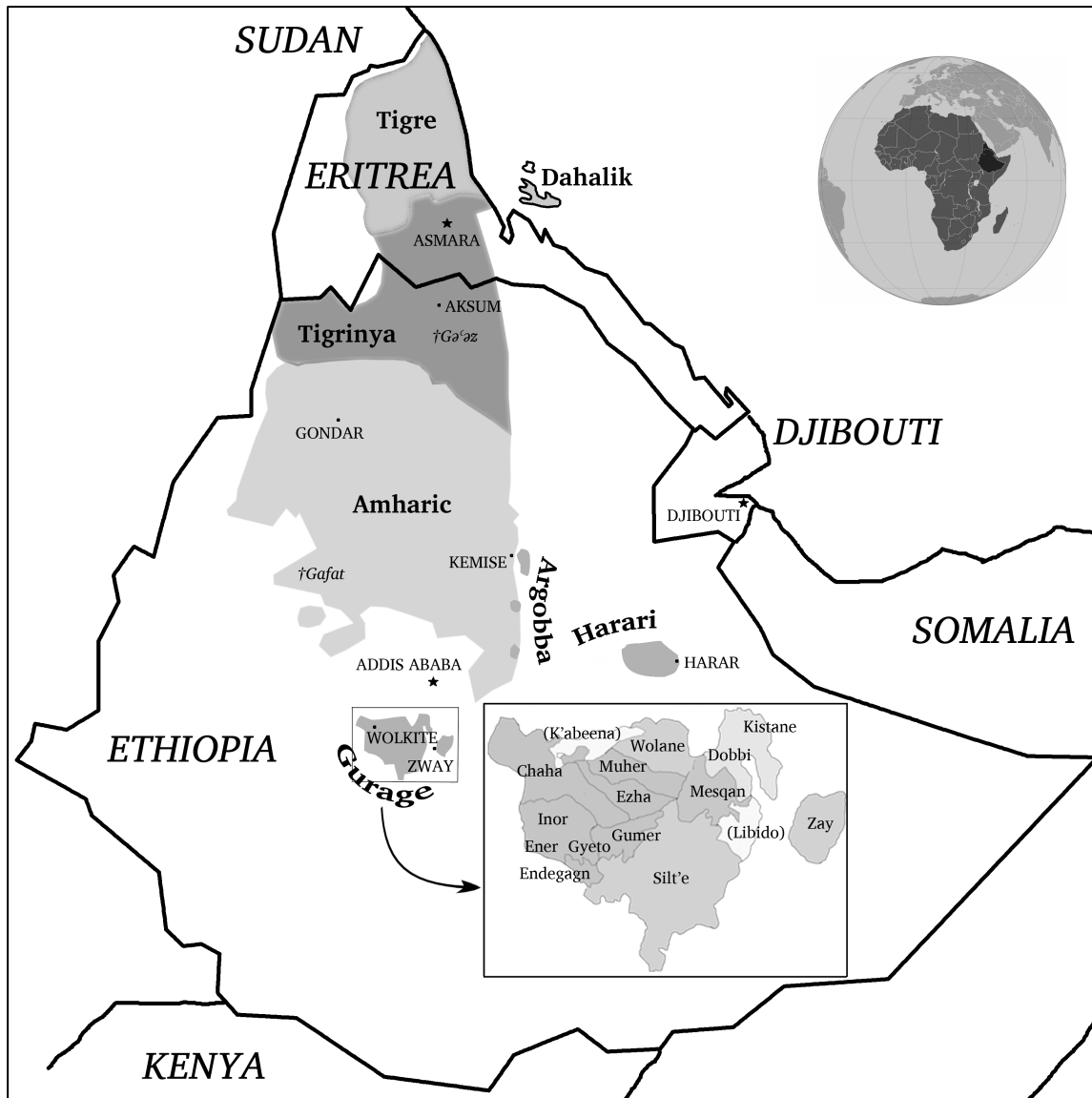


Figure 2: Geographical distribution of Ethiosemitic languages  
(taken from Meyer 2016a: 162).

The position of adverbs and adverbials is flexible. Time adverbs usually occur sentence initially, as in (1c), but they can also follow an overt subject noun or immediately precede the verb. Furthermore, adverbs can be marked for definiteness and accusative case to emphasize them, as the second occurrence of *zare* 'today' in (2).<sup>3</sup>

<sup>3</sup> The data quoted from *written* Amharic sources is given in the Amharic script (for which see Meyer 2016b) and in transliteration, whereas the data from oral sources are only presented in transcription.

(2) Source: Twitter<sup>4</sup>

ሸክ	ያለውን	ዛፍ	ዛሬ	ተከለገሁ
<i>fok</i>	<i>j-allä-w-n</i>	<i>zaf</i>	<i>zare</i>	<i>täklattfih<sup>w</sup></i>
thorn	REL-exist:NPST.SBJ.3SM-OBJ.3SM-ACC	tree	today	plant\CNV.2PL
ዛሬውን	ብትከት			
<i>zare-w-n</i>	<i>bi-tti-näku-t</i>			
<b>today-DEF.M-ACC</b>	<b>CND-SBJ.2PL-touch\PL-OBJ.3SM</b>			
አይወጋችሁም።				
<i>a-j-wäga-attfihu-mm.</i>				
	NEG-SBJ.3SM-pierce\IPFV-OBJ.2PL-NEG			

‘(If) you plant a tree with thorns [lit. a tree to which exist thorns] today and touch it the same day [lit. the today], it will not hurt [lit. pierce] you.’

Simple nouns can also become adverbs, but are then usually followed by a definite marker and the accusative suffix *-n*, as in the noun *ginbar* ‘forehead’ (15a).

As in other Semitic languages, verbs are formed through non-linear morphology by interdigitating a grammatical template with a lexical root. In other word classes (e.g. nouns, pronouns, interjections), fully vocalized stems predominate (for further details see, among others, Meyer 2011b; 2016c: 181–183; Edzard 2019; Mengistu 2023).

In verb inflection, Amharic has four primary conjugations, which are shown for the root √ s-b-r ‘break’ and the 3SM subject index in Table 1.

Conjugation	Template	Base	Index	Verb	
Perfective	C <sub>1</sub> äCC <sub>2</sub> äC <sub>3</sub>	*säbbär	-ä	säbbärä	‘he broke’
Imperfective	C <sub>1</sub> äC <sub>2</sub> C <sub>3</sub>	*säbr	j-	jisäbir	‘he breaks’ (dependent)
Imperative	C <sub>1</sub> C <sub>2</sub> äC <sub>3</sub>	*sbär	j-	jisbär	‘he should break’
Converb	C <sub>1</sub> äC <sub>2</sub> C <sub>3</sub> ä	*säbrä	-u	säbro	‘he breaking’ (dependent)

Table 1: Primary conjugations in Amharic.

The perfective and the imperfective encode two different viewpoint aspects, while the imperative is a mood expressing a command or request for the first, second and third persons. The fourth conjugation represents the converb (for details, see Meyer 2012).

<sup>4</sup> <https://twitter.com/natnaelmekonne7/status/1378704197667991557>, 07/01/2024

All four conjugations consist of a grammatical template and affixed person indexes that cross-reference the subject. The perfective and the converb index their subject with two different sets of suffixes, while the imperfective and the imperative use almost the same set consisting of prefixes and suffixes (see, e.g., Edzard 2019: 213). The pattern of the templates depends on the number of root consonants and the verb type (for further details, see Meyer 2011b: 1192–1193), i.e. the patterns in Table 1 are only valid for trilateral roots of Type A. Primary and applied objects can be indexed by a special suffix set, whereas the applied object index is preceded either by the malefactive/instrumental/locative (MIL) suffix *-bb*, as in (1c), or the benefactive suffix *-ll*. The imperfective verb is a dependent form which combines either with prefixes marking various types of subordinate clauses or with an auxiliary in affirmative main clauses. The converb, a verb form that cannot be negated, either marks a dependent clause or combines with a temporal auxiliary to form the perfect in main clauses. The most common auxiliaries are the cliticized existential verb = *all(ä)* for the nonpast and *näbbär* for the past. Negation is marked by the prefix *a(l)-* attached to the inflected verb and, in main clauses, additionally by the focus marker *-mm*, as in (2) (see Girma & Meyer 2008).

Amharic has a large number of ideophones (Goldenberg 2013: 223-224; Wetter 2003; 1999; Crass et al. 2001), but they have not received much attention in grammars. Amharic ideophones are morphologically invariant elements that usually occur only in combination with other morphemes. Most often, they are followed by the inflected verb ‘say’ with which they form a composite verb, such as the ideophone *t’äbb* ‘drop’ in (3).

(3) Source: Youtube<sup>5</sup>

ማር	ጠብ	ይላል፤	ካፍሽ።
<i>mar</i>	<i>t’äbb</i>	<i>ji-l = all</i>	<i>kaf-ij</i>
honey	IDPH.drop	3SM-say\IPFV = AUX.NPST	from.mouth-POSS.2SF
‘Honey drops from your mouth.’			

Following Mengistu (2010), I consider the composite verbs like *t’äbb jilall* ‘it drops’ in (3) to be complex predicates which consist of an invariant element, called coverb, which provides the lexical semantics and which is followed by a light verb that hosts

<sup>5</sup> <https://www.youtube.com/watch?v=r0r9ZyBFA5U>, 17/12/2023

the grammatical information. In (3), the coverb is the ideophone *t'äbb*, but other elements may also function as coverbs in Amharic, as will be shown in Section 4. The light verb is the verb *alä* 'say'<sup>6</sup> in the imperfective aspect marked with a 3SM subject. In complex predicates, the light verb has lost most of its original lexical semantic content and become an element that primarily encodes grammatical functions. Besides *alä*, there are a number of other light verbs in Amharic (see Section 4). Additional examples of ideophones with the light verb *alä* 'say' and the semantic fields they cover are shown in (4), which is based on Wetter (2003: 261–262).

(4) **Ideophones originating from imitative sound symbolism**

(= **onomatopoeia**)

- a. *t'wa*     *alä* 'pop, explode'
- b. *k'wa*     *alä* 'snap, click'
- c. *fiww*     *alä* 'whiz, swish, hiss'

**Ideophones originating from synesthetic sound symbolism**

(**depicting perceptible sensory imagery**)

- d. *billitt'*   *alä* 'flash, sparkle, glitter'
- e. *bikk'*     *alä* 'pop in, appear suddenly'
- f. *zimm*     *alä* 'be quiet'

**Ideophones originating from conventional sound symbolism**

(**depicting abstract concepts**)

- g. *zikk'*     *alä* 'be low/inferior'
- h. *k'äss*     *alä* 'be slow'
- i. *däss*     *alä* 'be pleasing'

Apart from Wetter (2003), the existing literature on Amharic has generally taken the complex predicate with 'say', with all its different coverbs, as a point of departure for analysis (Section 2).

In this paper, I will argue that Amharic ideophones form a distinct word class in Amharic that can be defined on language-internal morphosyntactic grounds. After a brief overview of the history of research on Amharic composite verbs with 'say' and related constructions (Section 2), I will discuss criteria that can be used to define

---

<sup>6</sup> As is common practice in the description of Semitic languages, the citation form of a verb is the 3SM perfective and not the infinitive/verbal noun.

ideophones as a word class in Amharic, which are typically realized as coverbs in a complex predicate (Section 3). Then I will discuss other coverbs that do not belong to the class of ideophones (Section 4). In addition to basic or simple ideophones, Amharic also has reduplicated ideophones and ideophones derived from a lexical root by templates, which are described in Section 5 and Section 6, respectively. To provide a broader picture of the distribution of complex predicates with the verb ‘say’ in Ethiosemitic languages, a comparative overview is presented in Section 7. A summary of the paper and the conclusions are given in Section 8.

The Amharic data in this paper comes from a convenience corpus consisting of various printed sources, primarily Leslau’s (1995) *Reference Grammar of Amharic*, my own observations of spoken Amharic over the past twenty years and targeted online searches. References have been provided for data cited from published sources and the Internet; unreferenced data are from observation. All data has been cross-checked with native speakers of Amharic (see Acknowledgements).

## 2. Research history

This section presents an overview of how complex predicates with ‘say’, which include ideophones, have been addressed in descriptions of Amharic over time.

The first Amharic grammar, Ludolf (1698a: 32), mentions a complex predicate with an ideophonic coverb, namely ገም ብኣሉኛ *zimm bi?aläh*<sup>w</sup> [sic!] (today *zimm al-k*<sup>w</sup> IDPH.silent say\PFV-1SG)<sup>7</sup> ‘I am silent’, without further explanation. In later grammars, the invariant coverb is usually treated as a particle or interjection combined with the verb *alä* ‘say’, or with the verbs *assäjñä* ‘feel to do; sense; name’, *täsäjñä* ‘be named’ or *adärrägä* ‘do, make’ (e.g. Isenberg 1842: 148–149; Praetorius 1879: 275–276; Armbruster 1908: 147).<sup>8</sup> A summary of early descriptions of these complex predicates can be found in Cohen (1939: 286–301). After an extensive list of examples, Cohen concludes that the coverb in these complex predicates is typically a “descriptive interjection” and proposes the term “composite descriptive verbs” to refer to them (Cohen 1939: 287). In post-World War II Amharic grammars, constructions with the verb ‘say’ have been called “composite verbs”, especially by Leslau (e.g. 1945a; 1995) and Appleyard (2001), but also by terms like “compound verbs” (e.g. Hetzron 1972:

<sup>7</sup> Note that the current form of the verb *alä* ‘say’ is historically related to the root b-h-l (see, e.g. Baye 1999) on which Ludolf’s expression is based.

<sup>8</sup> At least one Amharic grammar, Titov (1976), does not explicitly mention these constructions.

18; Baye 1999), which probably were then translated into Amharic as ጥምር ግሶች *t'immir gissottf* (lit. 'coalition verbs') (Baye 2008: 224–229), or “phrasal verbs” with the verb ‘say’ being a “supporting verb” (Meyer 2009: 27, see also Treis this volume for the Cushitic language Kambaata). Anbessa & Hudson (2007: 74–75) call these complex predicates “say verbs”.

Täklä Maryam (1964EC: 47), an early Amharic school grammar, considers the verb ‘say’ in complex predicates as an auxiliary that transforms nonverbal elements into verbs. This grammaticalization process is also assumed by Crass et al. (2001) and Wetter (1999). Mengistu (2010), whose terminology I have adopted in this paper, analyzes the verb *alä* ‘say’ and its causal counterpart *adärrägä* ‘do’ in complex predicates as “light verbs” and the preceding element as “coverb”. Note, however, that the term “coverb” is not limited to ideophones, but also includes other elements such as reported clauses, interjections, adverbials, simple nouns, reduplicated nouns and imperatives (see the examples in (14)).

The most comprehensive collection and detailed description of Amharic complex predicates with the verb *alä* ‘say’ and their coverbs can be found in Leslau (1995: 580–596). Baye (1999: 32–34) provides a semantic classification of all types of complex predicates with the verb ‘say’ in Amharic and thus implicitly also for ideophones. Based on the semantics and phonetic form of the coverb, he distinguishes between complex predicates for movement, silence, facial expressions, taste, rejection and onomatopoeic expressions.

Wetter (1999; and later 2002; 2003) was the first to use the term “ideophone” for the invariant coverb (i.e. non-reported element) in Amharic complex predicates with the verb ‘say’. In particular, Wetter (2003) elaborates on the concept of ideophone and identifies three types in Amharic, see (4), based on their degree of conventionalization by mapping their sound symbolic structure onto sensory imagery. Later, the term “ideophone” was also used in other work on Amharic, e.g. Zelealem (2011) and Meyer (2011b). The most recent publications on this topic are Teferi & Baye (2022a; 2022b), who present the preliminary results of an ongoing PhD study on Amharic ideophones and other elements occurring as coverb in constructions with the verb ‘say’ as well as various morphological and syntactic processes related to them.

In a broader context for Ethiopian languages in general, Appleyard (2001) proposes a form-oriented classification of complex predicates with the verb ‘say’ based on the morphological structure of the coverbs. He makes a very useful distinction between

lexical non-derived ideophones, quoted particles (mainly nouns and interjections), and coverbs derived from an existing verb root.

However, all these studies simply assume that some elements in combination with the verb 'say' should be considered as ideophones without providing a clear definition of what an ideophone actually is in Amharic or in other Ethiosemitic languages.

### 3. On the definition of ideophones in Amharic

Dingemanse (2012: 655) defines ideophones semantically as “marked words that depict sensory imagery”, which he proposes as a comparative concept for linguistic typology (for an overview on research on ideophones, see Dingemanse 2018). Morphologically and morphosyntactically, ideophones are usually determined by language-specific criteria, since common features that could universally define them have not yet been found (Voeltz & Kilian-Hatz 2001: 2; Kilian-Hatz 2020: 696). However, ideophones share a number of common tendencies (see Kilian-Hatz 2020 for details): They are usually sound-symbolic content-words with a language-specific semantics; they may exhibit phonological features not found in other word classes of the respective language; they are morphologically invariant and syntactically free; and they are preferably found in quotative constructions and used in informal speech situations.

In the following, it will be shown how these general phonological, morphological, morphosyntactic and semantic criteria apply to Amharic ideophones. The text extract in (5) illustrates the use of complex predicates with an ideophone as coverb in colloquial speech; the ideophones and their light verbs are given in bold.

#### (5) Source<sup>9</sup>

[Context: Several guests sit in a cafe bar drinking coffee and reading newspapers. Suddenly one of the guests jumps up frantically and starts searching for his cell phone. As he cannot find it, he accuses the waiter and other guests of theft. Eventually, the waiter asks him to pay the bill...]

---

<sup>9</sup> <https://www.sewasew.com/p/ፈታኛ፣ 17/01/2022>

- a. [...] የሂሳብ ደረሰኝ የመጣበትን ትንሽቶ  
*jä-hisab dāräsājn jä-mätt’a-bb-ät-in tinnif-ijje*  
 LNK-bill receipt LNK-come\PFV.SBJ.3SM-MIL-OBJ.3SM-ACC small-DIM  
 አቃፊ ነገር ብድግ ሲያደርገው —  
*ak’afi nägär biddigg s-j-adärg-äw*  
 holding thing IDPH.rise\_suddenly when-SBJ.3SM-do\IPFV-OBJ.3SM  
 ‘When he suddenly took up the rather small folder in which the bill came –’

- b. ከው!! እፍር!! ሽምቅቅ!!  
*kiww iffirr fimk’ikk’*  
 IDPH.very\_scared IDPH.very\_ashamed IDPH.be\_folded\_up  
 እንደው መግቢያ ጥፍት!!  
*indäw mägbija t’iffitt*  
 even entrance IDPH.absence  
 ‘He got frightened! He got ashamed! He pulled (his arms) over (his chest)! He did not know where to disappear!’

- c. «ምድር ዋጪኝ፤ ሰማይ ተከደንብኝ»  
*midir watf’i-jn sämaj tākädän-ibbi-jn*  
 earth swallow\IMP.SBJ.2SF-OBJ.1SG heaven cover\_up\IMP.SBJ.2SM-MIL-OBJ.1SG  
 አይነት ነገር፤  
*ajnät nägär*  
 kind thing  
 ‘(He was thinking) Something like “(Oh) earth swallow me, (oh) heaven cover me up” –’

- d. ሞባይሉ ቶብ አላለም!!  
*mobajl-u dubb al-al-ä-mm!*  
 cell\_phone-POSS.3SM IDPH.appear\_suddenly NEG-say\PFV-3SM-NEG  
 ‘(and then was it not that) his cell phone suddenly appeared (in front of him).’

The typical use of ideophones is illustrated in (5a) and (5d), where *biddigg* ‘rise suddenly’<sup>10</sup> and *dubb* ‘appear suddenly’ form a complex predicate with the verbs *alä*

<sup>10</sup> The word-middle consonant of some ideophones is optionally geminated or single, e.g. *biddigg* (Leslau 1995: 585) vs. *bidigg* (Kane 1990: 938) ‘rise suddenly’ (see also Cohen 1939: 288).



‘say’ and *adärrägä* ‘do’ as light verbs (see Section 4 below). The independent use of ideophones without a light verb is rare, but may occur in emphasized discourse (5b), in child-oriented speech, as in the extract from a children’s song in (6) and the riddle in (7), or in second-person imperatives (see e.g. Mengistu 2010: 299 fn. 4).

(6) Source: Youtube<sup>11</sup>

a. *wiħa-ye-n*                      *t’it’ t’itt~t’it’ t’itt.*

water-POSS.1SG-ACC IDPH.drink~PLR

‘(I) drink up my (glass of) water.’

b. *k’urs-e-n-imm*                      *bil’litt~bil’litt.*

breakfast-POSS.1SG-ACC-FOC IDPH.eat~PLR

‘And (I) also eat up my breakfast.’

(7) *liḡ-itwa*    *bikk’*    *t’il’likk’;*

child-DEF.F IDPH.pop IDPH.plunge

*inat-jä-wa*                      *k’oma*                      *dür’rikk’.*

mother-SNG-DEF/POSS.3SF stand\CNV.3SF IDPH.dry/insist

‘The child is [steadily] appearing and disappearing; her mother insists on standing still.’ [answer: pestle and mortar]

Although Amharic prosody is not yet well-studied, it is generally assumed that the last syllable of a polysyllabic word is not accented (Alemayehu 1987: 24, 29–31, 37; and also Leslau 1995: 44–45). For polysyllabic ideophones, on the other hand, it seems that the final syllable typically bears the accent, as shown by the symbol ‘ in (6) and (7). This observation still needs to be verified by further studies.<sup>12</sup>

With regard to their segmental phonemic structure and phonotactics, the majority of ideophones do not differ from other Amharic words. Only two ideophones—*nijñ alä* ‘buzz, hum’ and *jaw alä* ‘meow’<sup>13</sup> (Kane 1990: 1089)—begin with the palatal nasal, which usually does not occur word-initially in Amharic. The final consonant of an ideophone tends to be geminated. Word-final gemination, however, is not a

<sup>11</sup> <https://www.youtube.com/watch?v=BWyUIR0rpVY>, 17/01/2022

<sup>12</sup> Ideophones in the Bantu language Sena (seh) are also characterized by a different prosodic behavior vis-à-vis other word classes (see Guérois this volume).

<sup>13</sup> Or also *mijaw alä* (Kane 1990: 319) with a different initial nasal.

phonotactic peculiarity of ideophones and is also found with nouns (8) and verbs (9), as well as other word classes, e.g. the interjections *ettf* ‘ugh (expression of disgust)’ and *goff* ‘bravo’ (Leslau 1995: 907, 909), and bound morphemes such as the *-jn* POSS.1SG, *-ättf* 3SF subject index on perfective verbs.

(8) Leslau (1995: 12–13)

- a. *nättf* ‘white’ vs. *nättf* ‘one who plucks’
- b. *giff* ‘stripped’ vs. *gif* ‘wrong, injustice’
- c. *wätt* ‘solid, homogenous’ vs. *wät* ‘stew’

- (9) a. *t’iru-nna mät’fo säw-ottf-in läjj!*  
 good-and bad person-PL-ACC separate\IMP.2SM  
 ‘Distinguish between good and bad people!’
- b. *wiha-w jī-fäss näbbär.*  
 water-DEF.M 3SM-spill\IPFV AUX.PST  
 ‘The water was spilling over.’

It is instead a few interjections, not ideophones, that contain unusual sounds, as the paralinguistic clicks in (10), or the interjection *ärä* ‘really! (expressing surprise)’, the only word starting with the vowel *ä*.

(10) **Interjections with para-linguistic click consonants**

- a. Labial click  $(m)⊙^w$  (imitating a kissing sound)
- b. Alveolar click  $!!!$  (showing regret)
- c. Lateral click  $///$  (driving horses)
- d. Uvular click  $Ɔ Ɔ Ɔ$  (calling chicken)

The interjections with the click consonants in (10) are usually not mentioned in Amharic grammars and dictionaries (e.g. Leslau 1995: 899–909), probably because they are exclusively found in spoken language, and the Ethiopian script does not have graphemes to write them.

Although many ideophones in Amharic seem to be of onomatopoeic origin or to represent some kind of sound symbolism, e.g. *t’äbb* ‘drip’, *tʃ’ill~tʃ’ill* ‘tinkle’, *koff~koff*

'rustle' (Zelealem 2011: 3, 28, 31),<sup>14</sup> Amharic also has a number of regular nouns and verbs with an onomatopoeic sound pattern, as in (11).

(11) **Onomatopoeia**

<b>Noun</b>	<b>Verb</b>
a. Kane (1990: 1771, 1772) <i>dok'dokk'e</i> 'motorbike'	<i>tändäk'äddäk'ä</i> 'make a putt-putting sound'
b. Kane (1990: 2241) <i>tf'atf'ut</i> 'chick'	<i>täntf'attf'a</i> 'twitter'
c. Kane (1990: 1448) <i>kukkulu</i> (cock's crow)	<i>kokkolä</i> 'crow'

Consequently, an onomatopoeic sound pattern is not an appropriate criterion for defining an ideophone. Similarly, reduplication, which is often found in ideophones to express some kind of distributive meaning (see Section 5), also occurs in interjections, nouns, verbs, adverbs, etc.

Many ideophones can be nominalized by the suffix *-ta*, e.g. *billittf* (ideophone for flash, sparkle, glitter) > *billittf-ita* 'flicker (n), glimmer (n)', *zikk'* (ideophone for being low) > *zikk'-ita* 'lowness', *zimm* (ideophone for being quiet) > *zimm-ita* 'silence'. However, as the nominalization does not apply to all ideophones and also occurs with interjections, e.g. *imbi-ta* 'refusal' (< *imbi* 'no emphatic!'), *inde-ta* 'certainly, of course' (< *inde* 'what!, how!' expressing surprise), it cannot be considered a defining feature of the word class of ideophones.

Ideophones without a light verb appear obligatorily in the clause-final verbal slot, e.g. (6), or immediately preceding the light verb in complex predicates, as in (5a) and (5d). Their syntactic position is not flexible. Although ideophones without a light verb lack morphological person and TAM markers, they govern objects marked for the accusative case by the suffix *-n* (6) and can be modified by a converb (7). Thus, Amharic ideophones share syntactic features with verbs, not with nouns.

To recapitulate the main points: The ideophones formed by sound-symbolic imitation, e.g. (4a), are onomatopoeia, but not all onomatopoeia are ideophones. There is no phonetic or phonological feature that clearly separates them from

<sup>14</sup> Zelealem (2011) has a very broad concept of onomatopoeia. For example, he associates smallness with the vowel *i* and bigness with all non-front vowels, e.g. *t'ik'it* 'little (few)' vs. *bizu* 'many' (Zelealem 2011: 5), or the meaning of 'sucking' and 'absorbing' with the syllables *t'V* and *mVt'*, as in *t'äbba* 'suck' or *mätt'ät'ä* 'suck up, absorb' (Zelealem 2011: 11).

members of other word classes – apart maybe from an unusual word-final accent, which, however, needs further investigation. Ideophones share their obligatory clause-final position and their ability to govern direct objects with verbs, and their morphologically simple and invariant form with interjections. Moreover, interjections also occur in constructions with the verb ‘say’ (12a). In some of these constructions, the verb ‘say’ has grammaticalized into a light verb resulting in a complex predicate with a more abstract meaning, as in (12b), where the verb *alä* is no longer used to mark a direct quotation, but has become a light verb hosting grammatical information (see especially Wetter 2002).

(12) Wetter (2002: 1833)

- a. *kasa* « *imbi* » *al-ä*.  
 Kasa.M INTJ.no say\PFV]  
 ‘Kasa said, “no”.’
- b. *rädijo imbi al-ä*.  
 radio INTJ.no say\PFV-3SM  
 ‘The radio doesn’t work (lit. said “no”).’

In fact, most previous studies have treated all complex predicates based on the verb ‘say’ together, regardless of whether their coverb is an ideophone (4), an interjection (12) or some other element (see example (14) in Section 4). However, in these complex predicates, only ideophones, but not interjections or other elements as coverb, productively occur in noncausal/causal pairs with the light verbs ‘say’ and *adärrägä* ‘do’, respectively, as shown in (13). See also Cohen et al. (2002: 245).

(13) **Complex predicates with an ideophone as coverb**

	Noncausal		Causal	
a.	<i>t<sup>w</sup>a alä</i>	‘pop, explode’	<i>t<sup>w</sup>a adärrägä</i>	‘cause to pop’
b.	<i>billittf’ alä</i>	‘flash, glitter’	<i>billittf’ adärrägä</i>	‘cause to glitter, enlighten’
c.	<i>zikk’ alä</i>	‘be low/inferior’	<i>zikk’ adärrägä</i>	‘demote, degrade’

**‘say’ constructions with an interjection as coverb**

	Noncausal		Causal
d.	<i>imbi alä</i>	‘refuse’	* <i>imbi adärrägä</i>
e.	<i>mijaw alä</i>	‘meow’	* <i>mijaw adärrägä</i>
f.	<i>nijnj alä</i>	‘buzz, hum’	* <i>nijnj adärrägä</i>

In contrast to ideophones, interjections and other elements as coverb in (13) only cooccur with the verb ‘say’. While ideophones clearly differ from verbs and nouns in that they are morphologically invariant, the distinction between interjections and ideophones is akin to a grey zone in which the boundary between the two word-classes is not easily drawn, as in (12).

Based on these two features—typical occurrence as coverb and manifestation in a paired construction with the light verbs *alä* ‘say’ (noncausal) and *adärrägä* ‘do’ (causal)—ideophones can be considered a separate word class in Amharic. This word class differs from other Amharic word classes in that its members do not usually appear alone, but typically in combination with a light verb.

#### 4. The verb ‘say’ in combination with various coverbs

Constructions consisting of a coverb and the verb *alä* ‘say’ are very common in Amharic. The coverb can be made up of various elements: a quotation (14), a pseudo-quotation (14b), an ideophone (14c–d), a noun (14e–g),<sup>15</sup> a reduplicated imperative verb (14h)<sup>16</sup> and probably other elements (see also Goldenberg 2013):

- (14) a. « *nägä*      *addis abäbä*    *i-hed = allä-h<sup>w</sup>* »  
 tomorrow    Addis\_Ababa    1SG-go\IPFV = AUX.NPST-1SG  
*al-ä-jjñ.*  
 say\PFV-SBJ.3SM-OBJ.1SG  
 ‘He said to me, “I will go to Addis Ababa tomorrow.”’
- b. *bär-u*            « *ali-kkäffät* »            *al-ä-jjñ.*  
 door-DEF.M    NEG-1SG.be\_open\IPFV    say\PFV-SBJ.3SM-OBJ.1SG  
 ‘I could not open the door (lit. the door said to me, “I won’t open”).’
- c. *bär-u*            *g<sup>w</sup>a*            *blo*            *täzägga.*  
 door-DEF.M    IDPH.crash    say\CNV.3SM    be\_closed\PFV.3SM  
 ‘The door slammed shut.’

<sup>15</sup> For fully reduplicated nouns as coverb, see Section 6.

<sup>16</sup> Coverbs consisting of reduplicated imperative verbs occur invariably with a 2SM subject and typically describe an imagined action. This construction is only very rudimentarily described for Amharic and not found in the grammars of other Ethiosemitic languages.

- d. *mämt'at-u(-n)*                      *bät'am*    *däss*  
 come.VN-POSS.3SM-ACC    very    IDPH.happy  
*blo-ŋŋ = all.*  
 say\CNV.SBJ.3SM-OBJ.1SG = AUX.NPST  
 'His coming made me very happy.'
- e. Baye (1999: 35)  
***libb***    *blä-h*                      *tämälkät-äw!*  
 heart    say\CNV-2SM    observe\IMP.SBJ.2SM-OBJ.3SM  
 'Watch him/it carefully (lit. watch him/it saying, "Heart"!)
- f. *k-ajjä-h<sup>w</sup>-at*                              *wädih*    *fik'ir~fik'ir*  
 ABL-see\PFV-SBJ.1SG-OBJ.3SF    since    love  
*al-ä-ŋŋ.*  
 say\PFV-SBJ.3SM-OBJ.1SG  
 'Since I have seen her, I feel like being in love (lit. ..., [it] says to me, "love, love").'
- g. *wät'-wa*                      ***wiha~wiha***    *jät-l = all.*  
 stew-POSS.3SF    water                      3SM-say\IPFV = AUX.NPST  
 'Her stew tastes like water (lit. Her stew says, "water, water").'
- h. adapted from Baye (1999: 37)  
*s-aj-at*    ***inäk'-at~inäk'-at***  
 while-SBJ.1SG.see\IPFV-OBJ.3SF    strangle\IMP.SBJ.2SM-OBJ.3SF  
*al-ä-ŋŋ*  
 say\PFV-SBJ.3SM-OBJ.1SG  
 'When I see her, I very much feel like strangling her (lit. ..., [it] says to me, "Strangle her, Strangle her!").'

The expression *libb alä* 'pay attention' in (14e) is probably one of the many collocations with the noun *libb* 'heart' (see Kane 1990: 71–76) which eventually became a lexicalized idiomatic expression. Most typically the nouns acting as coverbs are repeated, as in (14f–g). All the coverbs in (14) have in common that no other free lexical element can intervene between the coverb and the light verb, see (21b) and (22b), and that the coverbs cannot be modified syntactically, e.g. by determiners or case markers, such as \**wiha-w~wiha-w al-ä* (water-DEF.M~water-DEF.M say\PFV-3SM)

or \**wiḥa~wiḥa-w-n alä* (water~water-DEF.M-ACC say\PFV-3SM) for the construction in (14g).<sup>17</sup> If a noun preceding the verb *alä* ‘say’ is marked for definiteness and accusative case (15a) or accompanied by a relational prefix (15b–c), it functions as an adverbial modification, not as a coverb.

(15) a. Baye (1999: 35)

*kasa aster-in ginbar-wa-n al-at.*  
Kasa.M Aster.F-ACC forehead-POSS.3SF-ACC say\PFV.SBJ.3SM-OBJ.3SF

‘Kasa hit Aster on her forehead (lit. Kasa said [it] to Aster on her forehead).’

b. Täklä Maryam (1964EC: 47, translation Ronny Meyer (RM))

ወባዋ እንደ መልቀቅ  
*wäba-wa indä- mä-lk’äk’*  
malaria-DEF.F like VN-leave\VN

ብላኛለኝ።

*bla-ገገ = allä-ttj.*

say\CNV.SBJ.3SF-OBJ.1SG = AUX.NPST-SBJ.3SF

‘I am almost cured from the malaria (lit. the malaria said to me [something] like leaving).’

c. *mata s-i-nnaddäd bä-boks*  
evening while-3SM-be\_angry\IPFV MIL-fist

*al-ä-ገገ.*

say\PFV-SBJ.3SM-OBJ.1SG

‘When he got angry in the evening, he hit me with a clenched fist (... he said [it] to me with a clenched fist).’

<sup>17</sup> The use of a reduplicated imperative verb as a coverb, as shown in (14h), deviates from this rule. The verb remains invariant in the 2SM, regardless of whether the speaker, who is the actual addressee of the imperative, is male or female. However, the object index is variable and must agree with the person, number and gender of the referent.

In contrast to the complex predicates with a coverb in (14), the adverbial does not necessarily immediately precede the verb *alä* ‘say’, but other constituents can intervene between them, as in (16), which are modifications of the sentences (15a, c).

- (16) a. *ginbar-wa-n*                      *sost*    *gize*    *al-at*.  
 forehead-POSS.3SF-ACC    three    time    say\PFV.SBJ.3SM-OBJ.3SF  
 ‘On her forehead, he hit her three times.’
- b. *bä-boks mata*                      *sost*    *gize*    *al-ä-jjɲ*.  
 MIL-fist    evening    three    time    say\PFV.SBJ.3SM-OBJ.1SG  
 ‘With a clenched fist, he hit me three times in the evening.’

The verb *alä*, originally meaning ‘say something to someone’, has different semantics depending on the constructions in which it occurs. In complex predicates with a coverb, it only indicates that the denoted situation is noncausal and serves as a host for verbal inflectional morphology (see below). However, when used with adverbials, such as in (15), it forms idiomatic expressions that convey the idea of ‘performing’ or ‘showing’ something related to the adverbial. Without an adverbial and in the imperative conjugation, *alä* functions as an encouragement to perform (usually to repeat) an action, or as a discourse particle with which the speaker announces the end of a discussion or his intended departure (see Meyer 2009: 32–33).

As seen in (14), the verb *alä* ‘say’ is morphologically irregular in Amharic. It uses different bases for the primary conjugations (see Table 1), i.e. *al-* (perfective, verbal noun), *l-* (imperfective), *blä-* (coverb) and *bäl-* (imperative), as well as the invariant base *bal* for the mediopassive (*tä-bal-ä* ‘be said’), indirect causative (*as-bal-ä* ‘let say’) and pluractional (e.g. the reciprocal *tä-ba~bal-ä* ‘say to each other’). Although these bases look like suppletive stems nowadays, they all developed from a single Ethiosemitic root, namely √ b-h-l ‘say’ (for details, see Baye 1999: 5–29; Leslau 1987: 89).

Semantically, the Amharic verb ‘say’ belongs to the group of verbs of speaking, some of which are shown in (17).

- |      |                                |                 |                                  |
|------|--------------------------------|-----------------|----------------------------------|
| (17) | INTRANSITIVE                   |                 | BITRANSITIVE                     |
| a.   | <i>tʃohä</i>                   | ‘cry’           | e. <i>t’äjjäk’ä</i> ‘ask’        |
| b.   | <i>tänʃokaffokä</i>            | ‘whisper’       | f. <i>azäzzä</i> ‘order’         |
| c.   | <i>täntäbattäbä</i>            | ‘stutter’       | g. <i>näggärä</i> ‘report, tell’ |
| d.   | <i>täk<sup>w</sup>älattäfä</i> | ‘stammer, lisp’ |                                  |



The intransitive verbs of speaking have only a subject argument, while the ditransitive verbs have two additional arguments: the addressee (indirect object) and the theme (direct object). The direct object can be an accusative-marked noun phrase (19) or an infinitive clause (18a), a complement clause marked by the complementizer *ind(ä)*- ‘that’ (18b), or a direct quotation marked by the converb of the verb ‘say’ (18c), which imbeds the quotations as theme object into the argument frame of the verb ‘ask’. Thus, the converb of ‘say’ in (18c) has a comparable function to the complementizer *indä*- in (18b) and is therefore also analyzed as part of the theme argument. Both objects of ‘say’ are optional, and each of them can also appear on its own.

- (18) a. *bä-sälam bet mä-gbat-ättfī-n*  
 [MIL-peace house VN-enter\VN-POSS.1PL-ACC]<sub>THEME</sub>  
*t’äjjäk’-ättf-in.*  
 ask\PFV-SBJ.3SF-[OBJ.1PL]<sub>ADDRESSEE</sub>
- b. *bä-sälam bet indä-gäbba-n*  
 [MIL-peace house COMP-enter\PFV-1PL]<sub>THEME</sub>  
*t’äjjäk’-ättf-in.*  
 ask\PFV-SBJ.3SF-[OBJ.1PL]<sub>ADDRESSEE</sub>
- c. *bä-sälam bet gäbba-tfih<sup>w</sup> bla*  
 [MIL-peace house enter\PFV-2PL say\CNV.3SF]<sub>THEME</sub>  
*t’äjjäk’-ättf-in.*  
 ask\PFV-SBJ.3SF-[OBJ.1PL]<sub>ADDRESSEE</sub>  
 ‘She asked us whether we got home safely.’

In (18), the addressee of the ditransitive verbs of speaking is marked by the 1PL object index *-n* on the verb *t’äjjäk’ä* ‘ask’. This object index is optional and can be omitted if the respective addressee is salient in the discourse, as in (19).

- (19) Source: Dictionary Abyssinica<sup>18</sup>

የቅኔ	ተማሪ	የቁጠረውን	ቅኔ
<i>jä-k’ine</i>	<i>tämari</i>	<i>jä-k’wätt’är-ä-w-in</i>	<i>k’ine</i>
LNK-poem	student	[LNK-count\PFV-SBJ.3SM-OBJ.3SM-ACC	poem] <sub>THEME</sub>
ለማሳረም	ለመምህሩ	ነገረ።	
<i>lä-massaräm</i>	<i>lä-mämhir-u</i>	<i>näggär-ä.</i>	

<sup>18</sup> <https://dictionary.abysinnica.com/ነገረ>, 23/01/2023

DAT-VN.correct\VN [DAT-teacher-POSS.3SM]<sub>ADDRESSEE</sub> report\PFV-3SM

‘The poetry student recited the poem which he had composed (lit. counted) to the teacher for correcting.’

All verbs of speaking can be modified adverbially by a coverb clause, which can be a construction consisting of a coverb and the verb ‘say’, as in (20).

(20) *ʔu~ʔu bla f’oh-ättf.*

IDPH.distress say\CNV.3SF cry\PFV-3SF

‘She cried out in distress.’

According to Mengistu (1996: 8–10; 1995: 4–6), the verb *alä* ‘say’ is a generic verb that lacks the inherent manner component of the other verbs of speaking in (17). Its immediately preceding coverb is not an object argument nor an adverbial modification, as it normally cannot be marked for definiteness and accusative case, be passivized (for further details, see Meyer 2009: 23–26) or be pronominalized (i.e. indexed by an object suffix on the verb). In Meyer (2009: 25), I argue therefore that the coverb and the light verb ‘say’ represent a fixed semantic and morphosyntactic construction in which the coverb does not function as a separate syntactic argument.

In contrast to the ditransitive verbs of speaking in (17), no syntactically free constituent can intervene between the coverb and the verb ‘say’, regardless of whether the coverb is an ideophone, a quotation, or another element. A construction analogous to (21a), in which the addressee (*lämämhiru* ‘to the teacher’) immediately precedes the verb *alä* ‘say’, is ungrammatical (21b).

(21) a. *jih mindin n-äw bla lä-mämhir-u t’äjjäk’-ättf.*

PRX.SM what COP-3SM say\CNV.3SF DAT-teacher-DEF.M ask\PFV-3SF

‘She asked the teacher saying, “What is this?”.’

b. *\*jih mindin n-äw lä-mämhir-u bla*

PRX.SM what COP-3SM DAT-teacher-DEF.M say\CNV.3SF

*t’äjjäk’-ättf.*

ask\PFV-3SF

Similarly, the degree adverb *bät’am* ‘very’ in (14d) cannot occur between the coverb and the verb ‘say’. Thus, a clause like (22b) is unacceptable; instead (22a) is used.

- (22) a. *bät'am däss blo-ገገ = all.*  
 very IDPH.happy say\CNV.SBJ.3SM-OBJ.1SG = AUX.NPST  
 'I am very happy (lit. it made me very happy).'
- b. \**däss bät'am blo-ገገ = all.*  
 IDPH.happy **very** say\CNV.SBJ.3SM-OBJ.1SG = AUX.NPST

However, bound grammatical morphemes can be prefixed to the verb 'say', as in (23a), and the contrastive focus marker *-mm* can be suffixed to coverbs (23b).

- (23) a. *t'ärmus-u sibbirr s-i-l*  
 bottle-DEF.M IDPH.**break\_completely** **while-3SM-say**\IPFV  
*bät'am dänäggät'-k<sup>w</sup>.*  
 very be\_shocked\PFV-1SG  
 'I was very shocked when the bottle broke completely.'
- b. Source: Sebiawi<sup>19</sup>
- የታገለጸ ለውጥ ያመጣል፤*  
*jä-tagääl-ä-w läwt' j-amät'all*  
 LNK-struggle\PFV-3SM-DEF.M change 3SM-bring\IPFV = AUX.NPST  
 'The one who struggles brings change.'
- ይልታገለጸ ቁጭ ብሎ*  
*j-al-tagääl-ä-w k'utt' blo*  
 LNK-NEG-struggle\PFV-3SM-DEF.M IDPH.sit\_down say\CNV.3SM  
*በባርነት ሲገዛ*  
*bä-barinnät s-i-ggäzza*  
 MIL-slavery while-3SM-be\_subjugated\IPFV  
*ይኖራል፤*  
*ji-nor = all;*  
 3SM-live\IPFV = AUX.NPST  
 'the one who does not struggle lives subjugated in slavery without any improvement (lit. by sitting down);'
- ቁጭም ባለበት*

<sup>19</sup> [http://sebiawi.blogspot.com/2014/01/blog-post\\_5253.html](http://sebiawi.blogspot.com/2014/01/blog-post_5253.html), 31/08/2023

<b><i>k'uttj'-imm</i></b>	<b><i>bä-al-ä-bb-ät</i></b>	
IDPH.sit_down-FOC	MIL-say\PFV-SBJ.3SM-MIL-OBJ.3SM	
ጾጥታል	ማለት	ነው።
<i>ji-mot = all</i>	<i>malät</i>	<i>n-äw.</i>
3SM-die\IPFV = AUX.NPST	say\VN	COP-3SM

‘that means that he dies in exactly the situation in which he is (lit. where he is sitting down).’

Many complex predicates with the verb ‘say’ seem only to have a subject argument, and thus appear to be intransitive (see, e.g., Leslau 1995: 583; and for Ethiosemitic in general, Cohen et al. 2002: 236; Ferguson 1970: 74). However, some of them also have an overt object (24a) or cross-reference it with an object index on the verb (24b) (see Meyer 2009: 36–38). Formally, the object constituents in (24) are marked like a primary object of a monotransitive verb, i.e. with the accusative suffix *-n* in (24a) and with the simple object index on the verb in (24a, b) – in contrast to the applied object in (25), in which the object index is preceded by the applicative suffix *-bb* (glossed MIL).

- (24) a. *hizb-u*                      *korona-n*  
 [people-DEF.M]<sub>i</sub>    [COVID-ACC]<sub>j</sub>  
*tfila*                      *blo-t = all.*  
 IDPH.ignore    say\CNV.SBJ.3SM<sub>i</sub>-OBJ.3SM<sub>j</sub> = AUX.NPST  
 ‘The people ignored the COVID disease.’
- b. *goräbet-e-n*                      *sälam* *s-i-l-at*  
 neighbor-POSS.1SG-ACC    peace    while-SBJ.1SG-say\IPFV-OBJ.3SF  
*zimm*                      *al-ättf-ijñ.*  
 IDPH.silent    say\PFV-SBJ.3SF-OBJ.1SG  
 ‘When I greeted my neighbor, she kept silent to me.’
- (25) *atkilt-ottf-e*                      *t'iwilligg* *al-u-bb-ijñ.*  
 vegetable-PL-POSS.1SG    IDPH.wilt    say\PFV-SBJ.3PL-MIL-OBJ.1SG  
 ‘My vegetables wilted on me [i.e. to my detriment].’

Complex predicates with the verb ‘say’ can also function as experiencer verbs, by cross-referencing the experiencer with the object index on the verb (26).



causal counterpart *adärrägä* ‘do’. Furthermore, there are also a number of complex predicates based on the verb *adärrägä* ‘do’, not *alä* ‘say’, whose coverb is often a loanword (see Section 7).

As with the verb *alä* ‘say’, the ideophones in complex predicates with *adärrägä* ‘do’ do not represent a theme argument or an adverbial, but are simply juxtaposed to the light verb with which they form a fixed lexical and morphosyntactic construction. As shown in (27), no other free syntactic constituent can intervene between the ideophone and the light verb *adärrägä* ‘do’. Sentence (27b), for instance, is unacceptable because the place adverbial *wädä gon* appears between the ideophone and the light verb.

- (27) a. *borsa-wa-in wädä gon k’uttj’ adärräg-ättj.*  
 bag-DEF.F-ACC towards side IDPH.sit do\PFV-3SF  
 ‘She put her bag down sideways.’
- b. \**borsa-wa-in k’uttj’ wädä gon adärräg-ättj.*  
 bag-DEF.F-ACC IDPH.sit towards side do\PFV-3SF

The light verb *adärrägä* ‘do’ is typically used (mono)transitively, but can also appear like an intransitive verb with a single subject argument, especially with ideophones expressing an activity, as in *f’iffirr adärrägä* (IDPH.dance do\PFV.3SM) ‘he does dancing’ (Mengistu 2010: 301). In transitive use, the valence of *adärrägä* includes a patient object that is directly affected by the verbal event.

Although the opposition noncausal/causal is typical for complex predicates with an ideophone as coverb, for semantic reasons not all ideophones occur with the two light verbs (Mengistu 2010: 301–303). For example, most ideophones derived from an intransitive root (see Section 6) whose agentive subject is not affected by the verbal event lack the noncausal complex predicate, as in *ziffinn adärrägä* ‘do singing’ (< √ z-f-n ‘sing’) without noncausal \**ziffinn alä*. In contrast, ideophones derived from intransitive roots in which the subject is also affected by the verbal event often lack the causal complex predicate, e.g. *adägg alä* ‘grow a little bit’ (< √ a-d-g ‘grow’), but no causal \**adägg adärrägä*.

Another peculiar feature of certain ideophones in complex predicates is their ability to combine with another pair of light verbs, namely noncausal *täsäñnä* ‘be named, be desired’ vs. causal *assäñnä* ‘name, feel like doing’, as in *däss täsäñnä* ‘be happy’ vs. *däss assäñnä* ‘make happy’ (Leslau 1995: 584; Kane 1990: 560–561). Whereas the occurrence



Full reduplication, however, is also found with other types of coverb in complex predicates (see, e.g., examples (14f–h)). Full reduplication occurs with all ideophones (31a–f), whereas partial reduplication only occurs with multisyllabic ideophones (31f–i).

- (31) Leslau (1995: 593–594)
- |    |                               |                                |
|----|-------------------------------|--------------------------------|
| a. | <i>t'äbb alä</i>              | 'drop'                         |
|    | <i>t'äbb~t'äbb alä</i>        | 'drip'                         |
| b. | <i>zikk' alä</i>              | 'be low'                       |
|    | <i>zikk'~zikk' alä</i>        | 'get continuously lower'       |
| c. | <i>tʃ'arr adärrägä</i>        | 'scratch a little or abruptly' |
|    | <i>tʃ'arr~tʃ'arr adärrägä</i> | 'scratch here and there'       |
| d. | <i>t'ät'a adärrägä</i>        | 'sip, drink slowly'            |
|    | <i>t'ät'a~t'ät'a adärrägä</i> | 'drink hurriedly'              |
| e. | <i>mullittʃ' alä</i>          | 'slip off'                     |
|    | <i>mullittʃ'~mullitʃ' alä</i> | 'become slippery'              |
| f. | <i>billittʃ' alä</i>          | 'flash suddenly'               |
|    | <i>billittʃ'~billitʃ' alä</i> | 'keep flashing continually'    |
|    | <i>bilittʃ'~littʃ' alä</i>    | 'glitter'                      |
| g. | <i>ziggitt adärrägä</i>       | 'close suddenly and tightly'   |
|    | <i>zigit~gitt adärrägä</i>    | 'close all around'             |
| h. | <i>bissitt adärrägä</i>       | 'pierce completely'            |
|    | <i>bisit~sitt adärrägä</i>    | 'riddle (with bullets)'        |
| i. | <i>nikk'il alä</i>            | 'be suddenly uprooted'         |
|    | <i>nik'il~k'ill alä</i>       | 'be completely uprooted'       |

A formal distinction between full and partial reduplication is found in the gemination pattern. While in full reduplication the gemination of consonants is retained (31a–e), in partial reduplication it is limited to the final consonant, as shown in (31g–i) and (32b).

Partial and full reduplication of ideophones has a number of senses that can be derived from a general distributive or pluractional meaning, namely senses of higher intensity (generally continuity and frequency of the denoted verbal event), as in (31a, b, e–i), but also of accelerated speed (31d) (Leslau 1995: 593). In addition, Baye (2008: 226–227) notes that multiple partial reduplication is used to express emphasis, e.g. regarding the speed of the action or the commitment of the subject, as in (32).





(34) Leslau (1995: 594–596)

- |    |                             |                 |  |
|----|-----------------------------|-----------------|--|
| a. | <i>k'uttf' biddigg</i>      | <i>alä</i>      | 'be restless'                          |
|    | (lit. sit_down + stand_up)  |                 |  |
| b. | <i>wät'a gäba</i>           | <i>alä</i>      | 'move to and fro'                      |
|    | (lit. go_out + go_in)       |                 |  |
| c. | <i>käff zigg</i>            | <i>alä</i>      | 'heave'                                |
|    | (lit. rise + be_lowered)    |                 |  |
| d. | <i>likk'imm t'irk'imm</i>   | <i>adärrägä</i> | 'clean up everything'                  |
|    | (lit. gather + hold_tight)  |                 |  |
| e. | <i>iffin t'iffinn</i>       | <i>adärrägä</i> | 'cover up completely to choke someone' |
|    | (lit. suffocate + cover_up) |                 |  |
| f. | <i>t'ikk mint'ikk</i>       | <i>alä</i>      | 'be very stubborn'                     |
|    | (lit. be_stubborn + ?)      |                 |  |
| g. | <i>kimbiss mimbiss</i>      | <i>alä</i>      | 'speak incoherently, nonsense'         |

While the origin of many of the echo-words can still be related to simple ideophones or verb roots, some of them, e.g. (34f, g), do not have an obvious relation to any other lexeme. As the echo-words occur with the verbs *alä* 'say' and *adärrägä* 'do' as light verbs, they are considered ideophones. Teferi & Baye (2022a: 106–108) call these constructions "compound ideophonic verbs".

However, Amharic also has a number of echo-words that are used like nominals and that never occur as coverbs, such as *arti murti* 'nonsense speech' and *tirki mirki* 'rubbish, empty chatter' (see Bezza 2013: 58–60; Teferi & Baye 2022a: 107). Thus, as is the case of onomatopoeia, not all echo-words are ideophones, even if the majority of them belong to this group.

## 6. Derived ideophones

In addition to the lexical word class of ideophones, Amharic also productively derives ideophones by interdigitating a lexical root with one of the two templates for trilateral roots:  $C_1CC_2CC_3$  for an intensive reading and  $C_1\check{a}C_2\check{a}C_3$  for an attenuated reading, compared to the event expressed by a regularly inflected verb from the same root. This derivation is shown in (35) with the trilateral root  $\sqrt{s_1-b_2-r_3}$  'break' (for other verb types and bi- and quadrilateral roots, see Leslau 1995: 586–593).

- (35) a. INTENSIVE TEMPLATE:  $C_1CC_2CC_3$   
           *sibbirr alä*           ‘break completely, suddenly (ITR)’  
 $s_1ibb_2irr_3$    *sibbirr~sibbirr alä* ‘(several things) break completely (ITR)’  
           *sibbirr adärrägä*   ‘break completely (TR)’
- b. ATTENUATIVE TEMPLATE:  $C_1äC_2äC_3$   
           *säbär alä*           ‘crack (ITR)’  
 $s_1äb_2är_3$    *säbär~säbär alä* ‘(several things) crack (ITR)’  
           *säbär adärrägä*   ‘crack (TR)’

Derived intensive and attenuative ideophones behave exactly like lexical ideophones, i.e. they appear as the coverb of the light verbs *alä* ‘say’ and *adärrägä* ‘do’, and they can be reduplicated. Compared to conjugated simple verbs, complex predicates with the intensive and attenuative templates denote a gradation in the intensity or force of the verbal event (see e.g. Leslau 1995: 582; Baye 1999: 30–31; Mengistu 2010: 296–297; Teferi & Baye 2022a: 108–110).

Since these derived ideophones are very productive in Amharic, most verbs occur in a tripartite system that distinguishes between the neutral form, i.e. the regularly conjugated verb, e.g. for (35), the noncausal mediopassive *täs<sub>1</sub>äbb<sub>2</sub>är<sub>3</sub>ä* ‘break ITR’ or the causal simplex *s<sub>1</sub>äbb<sub>2</sub>är<sub>3</sub>ä* ‘break TR’, and its intensive and attenuative forms in (35a) and (35b) (see Mengistu 2010: 296–297).

The derivation of ideophones from other words is typologically rare. Apart from Amharic (and other Ethiosemitic languages, see Section 7), it occurs in some Bantu languages, such as Cilubà (lua) (Kabuta 2001), Shona (sna) (Franck 2014), Tetela (tll) (Tassa 2001) and Sena (Guérois this volume), and also in Bulgarian (bul) (Kovatcheva 2014).

## 7. Complex predicates with the verb ‘say’ in Ethiosemitic

Complex predicates consisting of an invariant coverb followed by the verb ‘say’ occur in all Ethiosemitic languages, and are called “(descriptive) compound verbs” or “composite verbs” in the literature. Most likely, some of these coverbs also represent

ideophones.<sup>22</sup> But ideophones are generally not distinguished from other elements (e.g. interjections or adverbials) in older grammatical descriptions. Cohen (1939: 287)—the first summary of the composite verbs in Ethiosemitic—considers them to be a language contact phenomenon due to Cushitic influence. This assumption was widely accepted by others—including Leslau (1945a: 72), Hetzron (1972: 18), Tosco (2000: 346), Appleyard (2001: 9), Cohen et al. (2002: 244)—mostly because Asian Semitic languages lack such constructions.<sup>23</sup> Ferguson (1970: 73–74) considers them a defining feature of the Ethiopian Linguistic Area (for further details, see Crass & Meyer 2008: 232–233). However, Wetter (1999; 2002), Crass et al. (2001), Cohen et al. (2002) and Güldemann (2005) show that these complex predicates result from a grammaticalization process of the respective verb ‘say’, which is also found in other languages, even outside Africa (see also Mengistu 2010: 296).

One remarkable construction with the grammaticalized verb ‘say’ contains a fully reduplicated noun as coverb, as in (36), (37a) or (14f, g).

- (36) a. *tʃ’äw~tʃ’äw alä* (salt~salt) ‘taste salty’  
 b. *k’ibe~k’ibe alä* (butter~butter) ‘taste/smell like butter’  
 c. *tʃ’is~tʃ’is alä* (smoke~smoke) ‘smell like smoke’

- (37) a. *bird~bird al-ä-ḥḥ.*  
 CNV.cold~PLR say\PFV-SBJ.3SM-OBJ.1SG  
 ‘I have the chills.’ or ‘I feel cold.’  
 b. *bärräd-ä-ḥḥ.*  
 be(come)\_cold\PFV-SBJ.3SM-OBJ.1SG  
 ‘I feel cold.’

Based on a survey of their distribution in Ethiosemitic, these constructions are very widespread (for the sources, see fn.below). They have been reported for Tigrinya

---

<sup>22</sup> See also Treis (this volume) for Kambaata (ktb), a Cushitic language spoken in Ethiopia. Like Amharic, Kambaata ideophones form a distinct word class and typically occur as coverbs with the respective light verbs for ‘say’ and ‘do’, marking an intransitive/transitive contrast. Additionally, other elements can function as coverbs in Kambaata, and the border between ideophones and interjections is vague.

<sup>23</sup> In fact, similar complex predicates occur sporadically in Hebrew (heb), Babylonian Aramaic (tmr) and Neo-Aramaic (cld) (Goldenberg 2013: 224–225).

(North Ethiosemitic), Amharic and Silt’e (both Transversal Ethiosemitic), and Kistane, Muher, Gumer and Endegagn (all South Ethiosemitic). Even the little known and now extinct language Mesmes has at least one construction of this type, *k’ot’o~k’ot’o barä* ‘arrange in orderly manner (lit. say *k’ot’o k’ot’o*)’ (Ahland 2010: 111 line 7), where the reduplicated element is probably the noun ‘place’. This construction has not been mentioned in the grammars of Geez (North Ethiosemitic), and Argobba and Zay (both Transversal Ethiosemitic). In North Ethiosemitic, it seems to be limited to Tigrinya, as it does not exist in Tigre and Dahaalik. Harari (Transversal Ethiosemitic) also lacks it. However, Dahaalik and Harari have a construction in which a simple noun can be combined with the verb ‘say’, as in the Dahaalik example (38).

- (38) Dahaalik (Afroasiatic, Semitic; PC Marie-Claude Simeone-Senelle, July 2019)
- |              |              |
|--------------|--------------|
| <i>hibes</i> | <i>bi:la</i> |
| work.N       | say\PFV.3SM  |
| ‘he worked’  |              |

According to Beniam (2013: 150–152), the construction of simple noun plus verb ‘say’ is very common and productive in Harari, and often coexists with a regularly inflected verb from the same root. The meaning of such constructions with respect to the regular verbs is still not well described, but the coverbs are clearly nouns.

The combination of a reduplicated noun as coverb and the verb ‘say’ typically describe the perception of tastes and smells, e.g. (14g) and (36), besides a few other sensations, such as love in (14f) or the perception of temperature in (37a). However, in contrast to constructions with an ideophone, a fully reduplicated noun as a coverb cannot form a causal counterpart by exchanging the light verb ‘say’ with the verb ‘do’ (or the causative of ‘say’). Furthermore, reduplicated nouns or imperative verbs as coverbs do not denote a change in intensity, speed, etc., as is the case with fully reduplicated ideophones (Section 5), but only express a sensory experience related to the semantics of the respective simple nouns. In some cases, the semantics of such a complex predicate denotes nuances that cannot be expressed by a simple verb, as in (37), where only the construction with the reduplicated noun ‘cold’ in (a) has the sense of ‘having the chills’, from which probably the second sense ‘feeling cold’ developed, which is also denoted by the simple verb in (37b).

Loan verbs from languages with fully vocalized verb stems are often integrated into Amharic using a complex predicate based on the light verb *adärrägä* ‘do’. This is

because these stems cannot be easily converted into a consonantal root that would fit the conjugational templates. The loan verbs from English in (39) illustrate this phenomenon.

- (39) a. *tajb adärrägä* <English *type* ‘type (v.)’  
b. *kalär adärrägä* <English *color* ‘color (v.)’  
c. *kopi adärrägä* <English *copy* ‘copy (v.)’  
d. *särğäri ädärrägä* <English *surgery* ‘operate’

The noncausal counterpart of the complex predicates in (39) are formed by exchanging *adärrägä* ‘do’ with its mediopassive derivation *tädärrägä* ‘be done’, e.g. *tajb tädärrägä* ‘be typed’.<sup>24</sup> A few of the loanwords, usually those with three consonants, may also be converted into a root, and then be inflected like a simple verb of Type B, e.g. *täjjäbä* ‘type’ or *källärä* ‘color’ for the examples (a) and (b) in (39).

In the previous sections, it has been shown that Amharic has two different types of ideophones that occur as coverb in a complex predicate with the verb *alä* ‘say’, as summarized in (40).

- (40) a. Basic lexical ideophones, e.g. *zimm alä* ‘be quiet’  
b. Derived degree ideophones through combining a verb root with either an intensive or attenuative template, e.g. *sibbirr alä* ‘break completely/suddenly’ vs. *säbärr alä* ‘crack’

Not much is known about the actual distribution of the two types of ideophones identified for Amharic in (40) in other Ethiosemitic languages. The results of a survey based on published sources and a short questionnaire answered by specialists in some of the languages are given in Table 2.<sup>25</sup>

---

<sup>24</sup> In Zay (Transversal Ethiosemitic), ideophones and loanwords systematically use different pairs of light verbs to express the noncausal/causal relation. With ideophonic coverbs, the light verbs are *ba:l* ‘say’ and its causatives *?abäj* ‘cause to say’ or *?atbi:n* ‘facilitate to say’, whereas loanwords as coverbs cooccur with the light verbs *han* ‘become’ and *san* ‘do’ (Meyer 2005: 164–171; see also Meyer 2006: 816–817).

<sup>25</sup> The survey is based on information gathered from the following sources: Amharic (Leslau 1995: 580–596), Argobba (Wetter 2010: 148–150, 214–215), Dahaalik (PC with Marie-Claude Simeone-Senelle July 2019), Endegagn (Yohannes 2015: 138–145; PC with Yohannes Adigeh, July 2019), Gafat

Language	Basic lexical ideophones		ideophones from other roots	
	only ‘say’	noncausal/causal contrast	no pairs	intensive/attenuative pair
<b>North Ethiosemitic</b>				
†Geez		yes	?	?
Dahaalik		yes	no	no
Tigre		yes	yes	
Tigrinya		yes		yes
<b>Transversal Ethiosemitic</b>				
Amharic		yes		yes
Argobba		yes		yes
Harari		yes		yes
EG Silt’e		yes	yes	
EG Zay		yes	yes	
<b>South Ethiosemitic</b>				
GG Kistane		yes		yes
GG Muher		yes	yes	
GG Gumer		yes	yes	
GG Endegagn		yes	yes	
†Gafat	yes		?	?

(EG East Gurage, GG Gunnän Gurage)

**Table 2:** Distribution of ideophones in Ethiosemitic.

Table 2 shows that ideophones exist in all Ethiosemitic languages, even if the situation for the two extinct languages, Geez and Gafat, is not entirely clear. For these two languages, the coverb of complex predicates with the verb ‘say’ could also be an interjection, as assumed by Praetorius (1886: 156–157) for Geez in (41).

---

(Leslau 1945b: 88 further examples given on pp. 150, 153, 180), Geez (Praetorius 1886: 157), Gumer (Völlmin 2017: 171–173; PC with Sascha Völlmin, July 2019), Harari (Beniam 2013: 138–154), Kistane (PC with Bedilu Wakjira, July 2019), Muher (own fieldwork data, 2010), Silt’e (Gutt 1997: 933; PC with Rawda Siraj, July 2019), Tigre (Raz 1983: 66–67; PC with Saleh Idris, July 2019), Tigrinya (Tsehaye 1979: 108–111; PC with Shimelis Mazengia and Dagne Machew, July 2019) and Zay (Meyer 2005: 164–167).

(41) Geez (Afroasiatic, Semitic)

- |    |                 |              |   |
|----|-----------------|--------------|---|
| a. | <i>s'ätt'</i>   |              | 'shh (interjection used to call for silence)' |
| b. | <i>s'ätt'</i>   | <i>ji-be</i> | i. 'He said, "Shh!"' > ii. 'He kept quiet.'   |
|    | INTJ/IDPH.quiet | 3SM-say\PFV  |   |

If we assume that Geez possessed constructions like those in the Amharic example (7), the element *s'ätt'* in (41a) could also be an ideophone in an emphatic construction without the verb 'say'. For Geez, at least one pair with the noncausal/causal contrast was reported (42).

(42) Geez (Afroasiatic, Semitic; Praetorius 1886: 156–157; Leslau 1987: 12)

- |    |             |               |    |                |                  |
|----|-------------|---------------|----|----------------|------------------|
| a. | <i>?oho</i> | <i>bihl-ä</i> | b. | <i>?oho</i>    | <i>?a-bäl-ä</i>  |
|    | IDPH.yes    | say\PFV-3SM   |    | IDPH.yes       | CAUS-say\PFV-3SM |
|    | 'he agreed' |               |    | 'he persuaded' |                  |

The Geez examples (41) and (42) illustrate the general problem of distinguishing between ideophones and other word classes, especially interjections. Some of them may actually belong to two word classes, without a strict demarcation line, as in the Amharic example (12), where *imbi* 'no' is a quoted interjection in (12a), but also resembles an ideophone in the broader semantic sense of 'refuse' in (12b).

The most important morphosyntactic criterion for the definition of ideophones in Amharic is that they cooccur with a noncausal/causal light verb pair. This probably presents a more general feature of ideophones in Ethiosemitic. The noncausal light verb is the verb 'say', while its noncausal counterpart is either the light verb 'do' or the causative of 'say' in Ethiosemitic. However, the absence of the causal or, less frequently, also the noncausal light verb does not automatically mean that the coverb is not an ideophone. The noncausal/causal contrast expressed by the verbs *alä* 'say' and *adärrägä* 'do' systematically excludes certain semantic verb types. Without going into detail, with the exception of Gafat, all of the other Ethiosemitic languages in Table 2 have at least a few coverbs that occur in contrasting noncausal/causal pairs, and thus most probably represent ideophones.

Ideophones derived from roots occur in most Ethiosemitic languages (see also Appleyard 2001: 5). They are missing only in Dahaalik, and probably also in Geez and Gafat. In the remaining languages of Table 2, they occur in two distributions: In Tigre, East Gurage and Gunnän Gurage except Kistane, ideophones can be derived



from roots through different patterns, which do not contrast in terms of intensity/attenuation. Amharic, Argobba, Harari, Kistane and Tigrinya, on the other hand, regularly derive two types of ideophones from roots in order to express an intensive vs. attenuative contrast, which is, in turn, opposed to a semantically neutral, regularly inflected verb, as described for Amharic in Section 6 and for Tigrinya in (43), where only gemination or its absence expresses the contrast in degree.

(43) Tigrinya (Afroasiatic, Semitic; PC Shimelis Mazengia, July 2019)

a. **Neutral**

*säbir-u*

break\CNV-3SM

‘he broke it’

b. **Intensive**

*sibbir*      *?a-bil-u*

IDPH.break CAUS-say\CNV-3SM

‘he broke it completely’

c. **Attenuative**

*sibir*      *?a-bil-u*

IDPH.break CAUS-say\CNV-3SM

‘he broke it somewhat’

All Ethiosemitic languages which express the intensity/attenuation contrast through derived ideophones from the same verbal root seem to be in close contact with Amharic. Argobba and Tigrinya border the Amharic homeland, and the Kistane homeland is close to Addis Ababa. Harari has a very large diaspora in Addis Ababa and other Ethiopian cities (see Beniam 2013: 4), where Amharic is used as a lingua franca. Thus, contact with Amharic could have triggered the introduction of the intensive/attenuative contrast in these languages.

## 8. Conclusion

In this article, I have argued that ideophones form a separate word class in Amharic whose members commonly cooccur as a coverb in a complex predicate with the light verbs ‘say’ and ‘do’. All Ethiosemitic languages have at least a few lexical ideophones that can express this noncausal/causal contrast, even though the causal light verb may differ across the languages. Some ideophones can also be the input of a nominalization process with the suffix *-ta*. Syntactically, ideophones resemble verbs. Their position is fixed to the clause-final syntactic slot, usually occupied by verbs, and ideophones can also take direct objects and be modified by coverb clauses.

Ideophones are morphologically simple forms that do not inflect, and which can be reduplicated to express a pluractional meaning. Phonologically, ideophones are characterized by a stress on the ultimate syllable, but do not exhibit uncommon segmental features.

Ideophones are the only coverbs in complex predicates with the two light verbs ‘say’ and ‘do’ which express a noncausal/causal contrast. Other coverbs, such as interjections, adverbials, reduplicated nouns and verbs, and quotations, do not express this contrast.

Moreover, complex predicates with the light verbs ‘say’ and ‘do’ are so common in Amharic that they must be considered a separate inflectional verb class, contrasting with the regularly inflected verbs formed by nonconcatenative morphology. This verb class is characterized by reduced morphological complexity: it inflects only with the paradigms of two light verbs and can only contrast noncausal and causal events using distinct light verbs. Probably for to this reason, complex predicates with the verb ‘say’ are a very productive strategy to incorporate loanwords that do not easily fit into nonconcatenative inflectional morphology.

Finally, Amharic has a special type of ideophones derived from roots by two distinct templates, expressing the contrast between an intensified and an attenuated event. Derivational processes of ideophonization are rare from a typological point of view, but they also exist in Ethiosemitic languages in close geographical contact with Amharic, namely Tigrinya, Argobba, Kistane and Harari.

## **Acknowledgements**

This study of ideophones in Amharic was supported by a public grant overseen by the French National Research Agency (ANR) as part of the program “Investissement d’Avenir” (ANR-10-LABX-0083). It contributes to the IdEx Université de Paris (ANR-18-IDEX-0001). The analysis has benefited greatly from many fruitful discussions with members of the EFL research groups *Ideophones and Interjections in a Typological, Areal and Diachronic Perspective* (axe 3, GL1) and *Contact-Induced Morphosyntactic Change* (axe 2). I am also grateful for the helpful comments made during the *Workshop on Ideophones and Interjections* at the *Annual Meeting of the Societas Linguistica Europaea* from 21–24 August 2019 in Leipzig and the *9th Annual Conference and Annual Research Week of the College of Humanities, Language Studies, Journalism and Communication*, 10–

12 May 2022, Addis Ababa. I would like to extend my sincere thanks to Maria Bulakh, Dagnaw Mache, Fekede Menuta, Rawda Siraj, Saleh Idris, Shimelis Mazengia, Marie-Claude Simeone-Senelle, Sascha Völlmin and Yohannes Adigew for providing me with data and information on several Ethiosemitic languages. I also express my appreciation to Sitti Gragn Nasir, Seid Ahmed Ali and Delombera Negga for their patience in explaining and correcting the Amharic data. Finally, I thank the anonymous reviewers and the editors, especially Yvonne Treis, for their constructive suggestions.

## Abbreviations

1, 2, 3 = first, second, third person	F = feminine	PFV = perfective
ABL = ablative	FOC = focus	PL = plural
ACC = accusative	GG = Gunnän Gurage	PLR = pluractional
AUX = auxiliary	IDPH = ideophone	POSS = possessive
C = slot for a consonant in an inflectional template	IMP = imperative	PROG = progressive
CAUS = causative	INTJ = interjection	PRX = proximal demonstrative
CND = conditional	IPFV = imperfective	PST = past
CNV = converb	ITR = intransitive	REL = relative clause marker
COMP = complementizer	LNK = linker (genitive, relative verb)	SBJ = subject
COP = copula	M = masculine	SF = singular-feminine
DAT = dative	MIL = malefactive, instrumental, locative	SG = singular
DEF = definite	NEG = negation	SM = singular-masculine
DIM = diminutive	N = noun	SNG = singulative
EC = ethiopian calendar	NPST = non-past	TR = transitive
EG = east Gurage	OBJ = object	VN = verbal noun (infinitive)

## References

- Ahland, Michael. 2010. *Language death in Mesmes: a sociolinguistic and historical-comparative examination of a disappearing language*. Dallas TX: SIL International.
- Alemayehu Haile. 1987. Lexical stress in Amharic. *Journal of Ethiopian Studies* 20. 19–43.
- Anbessa Teferra & Grover Hudson. 2007. *Essentials of Amharic*. Köln: Rüdiger Köppe.

- Appleyard, David. 2001. The verb 'to say' as a "verb recycling" device in Ethiopian languages. In Andrzej Zaborski (ed.), *New data and new methods in Afroasiatic linguistics – Robert Hetzron in memoriam*, 1–11. Wiesbaden: Otto Harrassowitz.
- Armbruster, Carl Hubert. 1908. *Initia Amharica. An introduction to spoken Amharic. Part I: grammar*. Cambridge: Cambridge University Press.
- Baye Yimam. 1999. The verb 'to say' in Amharic. *Journal of Ethiopian Studies* 32(1). 1–50.
- Baye Yimam. 2008. *Jəʔamarijña səwasiw [Amharic Grammar]*. Addis Ababa: Elleni Publishing.
- Beniam Mitiku. 2013. *Harari: a descriptive grammar*. Doctoral dissertation, Addis Ababa University.
- Bezza Tesfaw Ayalew. 2013. *The submorphemic structure of Amharic: toward a phonosemantic analysis*. Doctoral dissertation, University of Illinois, Urbana-Champaign.
- Cohen, David & Marie-Claude Simeone-Senelle & Martine Vanhove. 2002. The grammaticalization of 'say' and 'do': an areal phenomenon in East Africa. In Tom Güldemann & Manfred von Roncador (eds.), *Reported discourse: a meeting ground for different linguistic domains*, 227–251. Amsterdam: John Benjamins.
- Cohen, Marcel. 1939. *Nouvelles études d'éthiopien méridional*. Paris: Ancienne Honoré Champion.
- Crass, Joachim, Barbara Dehnhard, Ronny Meyer & Andreas Wetter. 2001. Von 'sagen' zum Verbbildungsmorphem: Die Grammatikalisierung des Verbs 'sagen' einmal anders. *Afrikanistische Arbeitspapiere* 65. 129–141.
- Crass, Joachim & Ronny Meyer. 2008. Ethiopia. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 228–249. Cambridge: Cambridge University Press.
- Dingemanse, Mark. 2012. Advances in the cross-linguistic stud of ideophones. *Language and Linguistics Compass* 6(10). 654–672. <https://doi.org/10.1002/lnc3.361>.
- Dingemanse, Mark. 2018. Redrawing the margins of language: lessons from research on ideophones. *Glossa: A Journal of General Linguistics* 3(1). 1–30. [https://doi.org/DOI: https://doi.org/10.5334/gjgl.444](https://doi.org/DOI:https://doi.org/10.5334/gjgl.444).
- Eberhard, David M. & Gary F. Simons & Charles D. Fennig (eds.). 2022. *Ethnologue: languages of the world*. 25th edn. Dallas, TX: SIL International. Online version. <http://www.ethnologue.com>. (1 September, 2022)

- Edzard, Lutz. 2019. Amharic. In John Huehnergard & Na'ama Pat-El (eds.), *The Semitic languages*, 202–226. 2nd edn. London: Routledge.
- Ferguson, Charles A. 1970. The Ethiopian language area. *Journal of Ethiopian Studies* 8(2). 67–80.
- Franck, Genevieve E. 2014. *Ideophones in Manyika Shona: a descriptive analysis of ideophones and their function in Manyika (Bantu)*. BA Thesis, State University of New York.
- Girma Awgichew Demeke & Ronny Meyer. 2008. The enclitic *-mm* in Amharic: reassessment of a multifunctional morpheme. *Linguistics* 46(3). 607–628.
- Goldenberg, Gideon. 2013. *Semitic languages: features, structures, relations, processes*. Oxford: Oxford University Press.
- Guérois, Rozenn. 2024. Ideophones in Sena (Bantu, Mozambique). *Linguistic Typology at the Crossroads* 4(1): 244–289.
- Güldemann, Tom. 2005. Complex predicates based on generic auxiliaries as an areal feature in Northeast Africa. In Erhard F. K. Voeltz (ed.), *Studies in African Linguistic Typology*, 131–154. Amsterdam: John Benjamins.
- Gutt, Ernst-August. 1997. Concise grammar of Silt'e. In Eeva H. M. Gutt & Hussein Mohammed Mussa (eds.), *Silt'e-Amharic-English dictionary (with a concise grammar by Ernst-August Gutt)*, 896–957. Addis Ababa: Addis Ababa University Press.
- Hetzron, Robert. 1972. *Ethiopian Semitic: studies in classification*. Manchester: Manchester University Press.
- Hetzron, Robert. 1977. *The Gunnän-Gurage languages*. Napoli: Istituto Orientale di Napoli.
- Hudson, Grover. 2013. *Northeast African Semitic: lexical comparisons and analysis*. Wiesbaden: Otto Harrassowitz.
- Huehnergard, John & Na'ama Pat-El. 2019. Introduction to the Semitic languages and their history. In John Huehnergard & Na'ama Pat-El (eds.), *The Semitic languages*, 1–21. 2nd edn. London: Routledge.
- Isenberg, Charles W. 1842. *Grammar of the Amharic language*. London: Church Missionary Society.
- Kabuta, Ngo Semzara. 2001. Ideophones in Cilubà. In Erhard F. K. Voeltz & Christa Kilian-Hatz (eds.), *Ideophones*, 139–154. Amsterdam: John Benjamins.
- Kane, Thomas Leiper. 1990. *Amharic-English dictionary*. Wiesbaden: Otto Harrassowitz.

- Kilian-Hatz, Christa. 2020. Ideophones. In Rainer Vossen & Gerrit Dimmendaal (eds.), *The Oxford handbook of African languages*, 695–703. Oxford: Oxford University Press.
- Kovatcheva, Mira. 2014. An unrecognized class of words in Bulgarian and their word-formation. *Rivista de Linguistica* 26(2). 183–208.
- Leslau, Wolf. 1945a. The influence of Cushitic on the Semitic languages of Ethiopia: a problem of substratum. *Word* 1. 59–82.
- Leslau, Wolf. 1945b. *Gafat documents: records of a South Ethiopic language – grammar, text and comparative vocabulary*. New York, NY: American Oriental Society.
- Leslau, Wolf. 1961. Echo words in Ethiopic. *Annales d'Éthiopie* 4. 205–238.
- Leslau, Wolf. 1987. *Comparative dictionary of Gə'əz: Gə'əz-English / English-Gə'əz with an index of the Semitic roots*. Wiesbaden: Otto Harrassowitz.
- Leslau, Wolf. 1995. *Reference grammar of Amharic*. Wiesbaden: Otto Harrassowitz.
- Ludolf, Hiob. 1698a. *Grammatica Linguae Amharicae*. Frankfurt/Main: Prostat apud Johannem David Zunnerum.
- Ludolf, Hiob. 1698b. *Lexicon Amharico-Latinum*. Frankfurt/Main: Prostat apud Johannem David Zunnerum.
- Mengistu Amberber. 1995. Functional verbs in predicate formation: event-type hierarchy and grammaticalization. In Jocelyn Ahlers, Leela Bilmes, Joshua S. Guenter, Barbara A. Kaiser & Ju Namkung (eds.), *Proceedings of the twenty-first annual meeting of the Berkeley Linguistics Society, February 17-20, 1995 – general session and parasession on historical issues in sociolinguistics/social issues in historical linguistics*, 1–11. Berkeley, CA: Berkeley Linguistic Society.
- Mengistu Amberber. 1996. The transitivity of verbs of “saying” revisited. In José Camacho, Lina Choueiri & Maki Watanabe (eds.), *The proceedings of the fourteenth West Coast Conference in Formal Linguistics*, 1–15. Stanford, CA: CSLI Publications.
- Mengistu Amberber. 2010. The structure of the light verb construction in Amharic. In Brett Baker & Mark Harvey (eds.), *Complex predicates: cross-linguistic perspectives on event structure*, 291–318. Cambridge: Cambridge University Press.
- Mengistu Amberber. 2023. Amharic. In Ronny Meyer, Bedilu Wakjira & Zelealem Leyew (eds.), *The Oxford handbook of Ethiopian languages*, 414–442. Oxford: Oxford University Press.
- Meyer, Ronny. 2005. *Das Zay: Deskriptive Grammatik einer Ostguragesprache (Äthiosemitisch)*. Köln: Rüdiger Köppe.

- Meyer, Ronny. 2006. Lokale Varietäten des Zay – Ein Survey. In Gerald Heusing (ed.), *Sprach- und literaturwissenschaftliche Beiträge zum 16. Afrikanistentag, Leipzig, 25./26.9.2003*, 147–169. Berlin: LIT.
- Meyer, Ronny. 2009. The quotative verb in Ethiosemitic languages and Oromo. In Joachim Crass & Ronny Meyer (eds.), *Language contact and language change in Ethiopia*, 17–42. Köln: Rüdiger Köppe.
- Meyer, Ronny. 2011a. The role of Amharic as a national language and an African lingua franca. In Stefan Weninger (ed.), *The Semitic languages: an international handbook*, 1212–1220. Berlin: De Gruyter Mouton.
- Meyer, Ronny. 2011b. Amharic. In Stefan Weninger (ed.), *The Semitic languages: an international handbook*, 1178–1212. Berlin: De Gruyter Mouton.
- Meyer, Ronny. 2012. The converb in Amharic. In Domyenyk Eades (ed.), *Grammaticalization in Semitic*, 165–192. Oxford: Oxford University Press.
- Meyer, Ronny. 2016a. Aspect and tense in Ethiosemitic languages. In Lutz Edzard (ed.), *The morpho-syntactic and lexical encoding of tense and aspect in Semitic*, 159–239. Wiesbaden: Otto Harrassowitz.
- Meyer, Ronny. 2016b. The Ethiopic script: linguistic features and socio-cultural connotations. *Oslo Studies in Language* 8(1). 457–492.
- Meyer, Ronny. 2016c. Emergence of tense in Ethio-Semitic. In Ronny Meyer & Lutz Edzard (eds.), *Time in languages of the Horn of Africa*, 179–236. Wiesbaden: Otto Harrassowitz.
- Meyer, Ronny. 2018. On the internal classification of Ethiosemitic. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 168(1). 93–124.
- Praetorius, Franz. 1879. *Die amharische Sprache*. Halle: Verlag der Buchhandlung des Waisenhauses [Reprint: Hildesheim: Olms, 1970].
- Praetorius, Franz. 1886. *Äthiopische Grammatik mit Paradigmen, Litteratur, Chrestomathie und Glossar*. Karlsruhe: H. Reither.
- Raz, Shlomo. 1983. *Tigre grammar and texts*. Malibu: Undena Publications.
- Täklä Maryam Fantaye. 1964EC. ስነ-ጥበብ ለሥነ ጽሑፍ (*hohitā t'ibāb zäsīnā s'ihuf*) [*Way of wisdom for the art of writing*]. Addis Ababa: Bərhanəna Sälam Printing Press.
- Tassa, Obombe-Lukumbu. 2001. La formation des radicaux déidéphoniques et des idéophones déverbatifs en tetela (dialected ewango). In Erhard F. K. Voeltz & Christa Kilian-Hatz (eds.), *Ideophones*. Amsterdam: John Benjamins.
- Teferi Mulugeta & Baye Yimam. 2022a. The morphology of Amharic ideophones. *ሥነ ልሳን Zena-Lissan* 31(1).

- Teferi Mulugeta & Baye Yimam. 2022b. The syntax of Amharic ideophones. *Macrolinguistics and Microlinguistics* 3(1–2). 1–14. <https://doi.org/10.21744/mami.v3n1/2.28>.
- Titov, Evgeniï Grigor'evich. 1976. *The Amharic language*. Moscow: Nauka.
- Tosco, Mauro. 2000. Is there an “Ethiopian language area”? *Anthropological Linguistics* 42(3). 329–365.
- Treis, Yvonne. 2024. Ideophones in Kambaata (Cushitic): grammar, meaning and use. *Linguistic Typology at the Crossroads* 4(1): 147-200.
- Tsehaye Teferra. 1979. *Reference grammar of Tigrinya*. Washington, DC: Georgetown University, School of Languages and Linguistics.
- Voeltz, Erhard F. K. & Christa Kilian-Hatz. 2001. Introduction. In Erhard F. K. Voeltz & Christa Kilian-Hatz (eds.), *Ideophones*, 1–8. Amsterdam: John Benjamins.
- Völlmin, Sascha. 2017. *Towards a grammar of Gumer. Phonology and morphology of a Western Gurage variety*. Doctoral dissertation, University of Zürich.
- Wetter, Andreas. 1999. *Zur Entstehung eines Auxiliars im Amharischen—Grammatikalisierungsprozesse im nordäthiopischen Sprachraum*. Master Thesis, Johannes Gutenberg University, Mainz.
- Wetter, Andreas. 2002. A new analysis of the alə compound verbs in Amharic based on the grammaticalization theory. In Baye Yimam, Richard Pankhurst, David Chapple, Yonas Admassu, Alula Pankhurst, & Birhanu Tefera (eds.), *Proceedings of the 14th International Conference of Ethiopian Studies, 6–11 November 2000*, 1829–1839. Addis Ababa: Institute of Ethiopian Studies.
- Wetter, Andreas. 2003. Ideophones in Amharic. In Kézié K. Lébiakaza (ed.), *Actes du 3e Congrès Mondial de Linguistique Africaine, Lomé 2000*, 257–267. Köln: Rüdiger Köppe.
- Wetter, Andreas. 2010. *Das Argobba: Eine deskriptive Grammatik der Varietät von Shonke und T'ollaha (Zentraläthiopien)*. Köln: Rüdiger Köppe.
- Yohannes Adigeh. 2015. *Descriptive grammar of Endegagn*. Doctoral dissertation, Addis Ababa University.
- Zealelem Leyew. 2011. Onomatopoeic-ideophones in Ethiopian languages, with particular reference to Amharic. *Journal of Ethiopian Studies* 44. 1–46.

#### CONTACT

ronny.meyer@inalco.fr



# Ideophones in Sena (Bantu, Mozambique)

ROZENN GUÉROIS

CNRS, UMR 8135 LANGAGE, LANGUES ET CULTURES D'AFRIQUE

UNIVERSITY KWAZULU-NATAL

Submitted: 30/03/2023    Revised version: 12/02/2024

Accepted: 12/02/2024    Published: 12/06/2024



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

## Abstract

Based on a recently collected fieldwork corpus, this paper offers an overview of ideophones in Sena, a Bantu language spoken along the Lower Zambezi River in central Mozambique. By closely examining the different features (phonology, morphology, syntax and semantics) of Sena ideophones, this paper aims to identify the language-internal aspects of ideophones. The claim is that Sena ideophones are best treated as a distinctive word class whose members can be used in a variety of syntactic constructions. A crucial aspect of their syntactic properties is their ability to be used as holophrastic predicates.

**Keywords:** ideophones; word classes; holophrastic predication; Sena; Bantu.

## 1. Introduction

This paper provides a detailed description of ideophones in Sena, an Eastern Bantu language coded as seh (ISO 693-3), nucl1396 (Glottolog) and N44 in Guthrie's (1967-71) referential classification. It is spoken along the lower Zambezi valley in central Mozambique in the provinces of Sofalá, Tete, Zambézia and Manica. The number of speakers is estimated to be more than 1,600,000 speakers (Instituto Nacional de Estatísticas 2017).

There is a vast body of literature on ideophones, but despite important typological research (e.g. Hinton, Nichols & Ohala 1994; Voeltz & Kilian-Hatz 2001; Dingemanse

2012; Haiman 2018; Akita & Pardeshi 2019), there is no strict consensus on either the definition or the functions of ideophones across languages. It appears that crosslinguistically there are different profiles of languages when it comes to defining ideophones. Classically considered as “vivid representations[s] of an idea in sound” (Doke 1935: 118), they were later defined as “marked words that depict sensory imagery” (Dingemanse 2012: 655). Ideophones have iconic properties that convey depiction associated to sensory perceptions as well as motion, temporal unfolding, feelings and cognitive states (Kilian-Hatz 1999: 31–52; Dingemanse 2012: 661; Dingemanse et al. 2015: 607). Controversies exist in terms of word class affiliation, i.e. whether ideophones can be subsumed under (an)other word class(es) or constitute their own part of speech. In African studies, ideophones are either analysed as a subcategory of adverbs (see, e.g. Doke 1931: 221 for Shona, *sna*), or they are considered as a distinct word class (Cole 1955).

By closely examining the different features (phonology, morphology, syntax and semantics) of Sena ideophones, this paper aims to offer a first analysis of this type of lexemes in the language, and reach a language-internal definition of ideophones. I argue that Sena ideophones are better treated as a separate word class, featuring phonological, morphological and syntactic properties not otherwise attested in other word classes.

The article is organized as follows. The basic morphosyntactic aspects of Sena word classes are presented in Section 2. Section 3 provides main properties of Sena ideophones as well as the inventory used in this study. Sections 4 to 7 examine their properties in terms of phonology, morphology, syntax and semantics, respectively. Finally, Section 8 offers some discussion and conclusions.

The data used in this paper come from Prieto (2015), which includes a dictionary section in which a fairly large number of ideophones were found, and two periods of fieldwork carried out by the author in Maputo (fall 2021) and in the town of Caia and its surroundings, in the north of the Sofalá Province (summer 2022). Throughout the paper, I use a phonological notation for the linguistic data. Note, however, that <y> stands for the palatal approximant /j/, as conventionally practiced in Bantu studies.

## **2. Major word classes in Sena**

In many respects, Sena is a typical Bantu language: its morphology is highly agglutinative and ruled by a gender system of agreement classes, and it follows a basic

SVO word order.<sup>1</sup> The canonical syllabic structure is CV, where C stands for any consonant, including prenasalized consonants, affricates, glides, and labialized/palatalized consonants. By default, the nucleus is the vowel, but syllabic nasals commonly appear noun-initially as a result of vowel apocope between a nasal onset and a following labial or coronal consonant, followed by homorganic nasal assimilation (e.g. noun class prefix 3 *mu-* > *ŋ-* in *ŋsoro* ‘head’). There are no closed-syllable words in Sena. The only exception to this generalization is ideophones, as examined in detail in §4.2. Unlike most Bantu languages, Sena does not have phonemic tones. It instead shows a rather predictable accentual system associated to the penultimate syllable, similar to that found in Swahili (swh).

The rest of this section briefly describes major word classes in Sena, i.e. word classes which have inflecting members, ruled by a predominantly head-marking morphology characterized by an extensive system of noun class agreement (gender-number agreement patterns). These are nouns, adnominal modifiers and verbs. Adverbs are also presented at the end of the section.

Sena nouns are generally bi- or trisyllabic. As a typical Bantu language, it has an elaborate system of noun classes (gender-number agreement patterns) numbered from 1 to 18. Structurally, nouns minimally consist of a stem to which a noun class prefix is assigned. Most classes function in singular/plural pairs. Compare, e.g. singular class 1 *ŋ-kazi* ‘woman’ versus plural class 2 *a-kazi* ‘women’. Class 15 *ku-* serves to form infinitives (e.g. *ku-ronga* ‘to speak’), hence it is glossed as INF throughout. The language also retained the three historical Proto-Bantu locative classes \*pà (class 16), \*kù (class 17), and \*mù (class 18) (Meeussen 1967; Grégoire 1975), realized as *pa-*, *ku-* and *mu-*, respectively (e.g. *pa-nyumba* ‘at home’). In the NP, nouns by default appear phrase-initially and function as the head of the construction. In verbal clauses, they usually serve as subjects and/or objects. There is no different marking of nouns in different argument positions or different oblique roles.

Nouns as heads of NPs may combine with adnominal modifiers, which usually follow them and agree with them through class prefixes. Adnominal modifiers include adjectives, demonstratives, possessives, numerals/quantifiers, and interrogatives. Some of these modifiers share the same agreement paradigm, others have their own.

---

<sup>1</sup> For more information on what typical Bantu languages are, the reader is referred to general works such as Williamson and Blench (2000: 11–42), Nurse and Philippson (2003) and Van de Velde et al. (2019).

In this paper, for ease of reading, agreement prefixes within noun phrases are all glossed as AGRP for ‘agreement prefix’, then followed by the noun class number.<sup>2</sup>

Like most Bantu languages, Sena has rich agglutinative verbal morphology. The verbal complex consists of a string of affixes and clitics assigned to specific verbal slots that are ordered around the verb root contained in the radical slot. As can be seen in Table 1, a range of grammatical information is conveyed by these elements.

Slots	pre-initial	initial	post-initial	pre-radical	radical	pre-final	final	post-final
Functions	TAM negation	subject	TAM AM	object	<b>verb</b> <b>root</b>	derivation TA	TAM	PLA locative subject pronouns in relatives, etc.

**Table 1:** The morphological structure of the finite verb in Sena  
(based on Güldemann 1999’s terminology).

The example in (1) illustrates the ordering and type of information that may be encoded in the verb form, with temporal and aspectual information (present *na-* prefix), subject and object agreement (*ndi-*, *ku-*), and derivational suffixes (applicative *-ir*).

- (1) *ine ndinakut<sup>h</sup>amangira*  
*ine ndi-na-ku-t<sup>h</sup>amang-ir-a*  
 PRO1SG SP1SG-PRS-OP2SG-run-APPL-FV  
 ‘I’m going to chase you away.’ (Story\_HareCobra\_#31)

Like a typical Bantu language, Sena displays nominative-accusative alignment. The unmarked and default word order is Subject-Verb(-Object)(-Adjunct/Oblique). The initial agreement slot of the conjugated verb form invariably expresses agreement with the subject (except in relative constructions). NP arguments whose identity is recoverable from the context can freely be omitted. This is especially true for subjects, which are indexed through verb agreement anyway.

<sup>2</sup> Readers are referred to Güldemann & Fiedler (2019) for an overview on noun classes and gender in Bantu and beyond.

Adverbs, along with nouns and verbs, are an open word class. In particular, the derivational prefix *ci-* (class 7) is a productive way of creating adverbs, especially manner adverbs. Most of them include reduplication (e.g. *cinjala-nyala* ‘always with hunger’ < *nyala* ‘hunger’), but non-reduplicated *ci-* adverbs are also attested (e.g. *cimpira* ‘in the shape of a ball’ < *mpira* ‘ball’). Several temporal adverbs are also based on the class prefix *ci-*, e.g. *cino* ‘now’, *cipo* ‘never’, *citi-citi* ‘midnight’, *cit<sup>h</sup>anjkwí* ‘at the beginning, formerly’, etc., however the lexical (or grammatical) sources from which they derive are synchronically not retrievable. Other temporal adverbs are uninflected, e.g. *rero* ‘today’, *mangwana* ‘tomorrow’ or *kare* ‘formerly’. As for locative adverbs, they usually take one of the three locative prefixes, e.g. *pa-kati/ku-kati/ṅ-kati* ‘in the middle’.

### 3. Sena ideophones: main characteristics and inventory

Ideophones are “marked words” (Dingemanse 2012: 654) in the sense that they stand out from other lexemes. How markedness manifests itself may differ from language to language and has to be defined in language-internal terms. This section summarizes the main characteristics of this word class in Sena (§3.1) and provides basic information on the inventory used in this study (§3.2).

#### 3.1 Main characteristics

The Sena language has been the subject of a number of studies led mostly by missionaries since the late 19<sup>th</sup> century. Anderson’s (1897) short grammatical account does not discuss ideophones at all. Torrend (1900: 173–176) refers to ideophones as “*informal words*” which can be used as *adjectives* (after *li* ‘be’), *adverbs* or *interjections*. Only a few examples are provided for each use, and none are accompanied by additional comments or analysis. There is no subsection on ideophones in Moreira (1924). A couple of ideophones can be identified as part of superlative constructions (p. 51). The author refers to them as “*invariable nouns*”. A few other ideophones are found in temporal expressions, such as *dzua pióo* ‘at sun rise’ (p. 159) or *dzua tobi* ‘after sunset’ (p. 160). Alves (1939: 40–41) talks about *adverbs* and proposes a handful of examples. The studies by Funnell (2004) and Kiso (2012) focus on some specific

aspects of Sena grammar. None describes ideophonic words. Finally, Prieto (2015) is divided into several sections, including a Sena-Portuguese lexicon and a short grammatical sketch. In the latter, ideophones are treated in one paragraph, where the author refers to them with the label “*onomatopoeic adverbial expression*” (Prieto 2015: 413). Here too, a handful of examples are given. Crucially, even if Prieto’s grammatical analysis of ideophones is very superficial, his Sena-Portuguese lexicon (p. 3–234) has the advantage of providing a significant number of ideophones (at least 400), which serve as the main source for the current study.

The way these different authors classify ideophones is unfortunately not based on any given definition. The variety of attributes provided and italicized above shows to what extent the authors disagree about their status. The rest of this section is devoted to the main characteristics of Sena ideophones. Only high-level points are highlighted here. Details are provided in the different following sections.

With regard to phonology (§4), ideophones in Sena only contain phonemes that also belong to the general phoneme inventory of the language. The most salient phonological properties of ideophones are the following: i) they recurrently resort to vowel lengthening (in a system where long vowels are non-phonemic); ii) they exhibit several strategies of reduplication; iii) they include a handful of items with closed syllables (CVC pattern), a syllabic structure unattested in other parts of speech; and iv) in a prosodic system marked by stress, ideophones exhibit tones.

With regard to morphology (§5), ideophones are uninflected, and, with a few exceptions, deprived of derivational morphology. Ideophones thus have the appearance of being underived roots. A frequent morphological process, however, involves the reduplication, triplication or multiplication of the radical element, something which is a common characteristic of ideophones cross-linguistically (see, e.g. Voeltz & Kilian-Hatz 2001: 2; Childs 2003: 8; Dingemans 2012: 656; Andrason 2020: 127–128). Many Sena ideophones which describe a property or manner have the shape of verb roots. The relationship between verbs and ideophones has long been postulated in Bantu (Doke 1954; Fortune 1961; Childs 1994). In Sena, the presence of derivational morphology in a handful of ideophones suggests a one-way derivational mechanism, that is deverbal derivation. Deverbal ideophones in Sena could be analyzed as defective verbs, but since they share the same properties as the other ideophones, they are treated as ideophones.

With regard to syntax (§6), Sena ideophones can appear in three different constructions summarized in (2).

- (2) - Complement of cognate verbs (adverbial use)
- Complement of copula verbs or light verbs (complex predicates)
- Ideophonic predicates (predicative and holophrastic use)

This wide distribution results in different degrees of syntactic integration. When they combine with cognate verbs, they behave as manner adverbs co-expressing the event expressed by the verb, such as ‘he fell IDEO(falling)’. They form complex predicates when they combine with copula verbs (‘be’ + IDEO) or light verbs (‘do’ + IDEO). When they are used predicatively and holophrastically, Sena ideophones function as “complete autonomous utterance[s]” (Andrason 2020: 130). They can substitute for verbs and function on their own as predicates with the capacity to induce argument positions. This corresponds to the prototypical syntactic definition of ideophones (Childs 1994; Dingemanse 2012: 656–657; Andrason 2020: 130).

Finally, with regards to semantics (§7), all semantic fields identified by Dingemanse (2012: 663), i.e. movements, all types of sensory perceptions, inner feelings and cognitive states, are covered by Sena ideophones. A complementary pattern ruled by phono-semantic principles is observed in many deverbal ideophones. The first pattern correlates vowel lengthening with a single, durative and usually intense event (e.g. *nyá:nygu* ‘licking’), while the second pattern semantically associates reduplication (or multiplication) with a recurrent, repetitive or rhythmic event (e.g. *nyangú-nyangú* ‘licking repeatedly’).

### **3.2 The inventory**

Bantu languages differ considerably as to their number of ideophones. Large inventories are posited for Duala (dua, Cameroon, Meinhof 1912), Nkundo (lol, North-Western DRC, Hulstaert 1938), and a number of central Bantu languages such as Tumbuka (tum, Malawi, Tanzania, Zambia, Young 1932). At the other end of the spectrum, Rundi (run, Burundi, Meeussen 1959) and Gusii (guz, Kenya, Whiteley 1956) are two Bantu languages with seemingly few ideophones. In between, there are languages like Tswana (tsn), where “ideophones do appear to be fewer in number and to be less frequently used in Tswana than in many other Bantu languages” (Cole 1955: 370). Unfortunately, speaking of large or small inventories does not say anything about exact numbers, and as can be expected, the differences in size assessment may easily be attributed to different definitions of the authors and/or the depth at which

ideophones have been described in each language. A dedicated study is needed to avoid these pitfalls.

Earlier work on Sena (§3.1) does not convey a precise idea of the number of ideophones and even less of their frequency in the language. In terms of numbers, the 412 lexemes found in Prieto probably represent just a sample. Prieto himself states that “[t]here is a whole dictionary of onomatopoeic and alike sentences to be written yet.”<sup>3</sup>

As for my own fieldwork corpus, excluding elicitation, a total of almost 6:30 hours of naturalistic data were recorded and transcribed.<sup>4</sup> These include radio broadcasts, conversations, life stories, folktales and descriptions of cultural practices. I also included poems written in the 1970s by two Sena speakers while they were living and studying in Zobwe, Tete Province. In Table 2, the different audio files are listed by genre. The number of recordings (with an indication of their total length in time) and the number of types and tokens are specified for each genre. In total, across the different genres, there are 175 instances (tokens) of ideophones for 54 types.

Genre	Number of recordings (+ recording time)	Type number of ideophones	Token number of ideophones
radio broadcasts	2 recordings (00:33:36)	1	1
life stories	11 recordings (03:03:50)	6	14
descriptions	9 recordings (00:50:47)	7	23
conversations	4 recordings (00:36:05)	5	37
folktales	12 recordings (01:02:33)	23	81
poems	11 recordings (00:21:40)	16	19

**Table 2:** Occurrences of ideophones (in types and tokens) based on available discourse genres.  
(Guérois 2021 – 2022 field data)

Among the discourse genres listed in Table 2, radio broadcasts and life stories have very few ideophones overall. The discourse genre ‘descriptions’ closely follows, with at most one or two ideophones for recordings between 5 and 15 minutes. Two very short recordings (between 01:30 and 2:00 minutes) have 10 and 7 ideophone tokens, respectively, but with 15 occurrences of *gedé* ‘completing, concluding an action’. The last three discourse genres in Table 2, conversations, folktales and poems, have

<sup>3</sup> Translated from the original: «Há um dicionário inteiro, ainda por escrever, de frases onomatopoeicas e similares.»

<sup>4</sup> As the raw data is still being processed, this corpus has not yet been made available in open access.



proportionally more ideophones. That conversations and folktales contain a considerable number of ideophones is not surprising, considering that ideophones are used for dramatic enhancement through sound effects. The speaker not only narrates the event but also dramatizes for his audience (or interlocutor) by the means of an ideophone (the linguistic feature), which may be accompanied by a simultaneous imitation, in the form of a gesture (the extralinguistic feature). It is also not surprising to find ideophones in poems, as this genre not only embraces orality through rhythm, sonic qualities and acoustic resonance, but also invokes imagery, just as ideophones typically do. Four ideophones extracted from poems are used in this study.

Another interesting fact, which would deserve more in-depth research, is speaker variation. The current corpus suggests that monolingual Sena speakers tend to use ideophones more frequently than non-monolingual speakers. Coincidentally, monolingual Sena speakers are more likely to live in rural areas and usually have a lower level of scholarly instruction.

The inventory of ideophones used in the present study brings together the 54 types of ideophones found in fieldwork data and those provided in Prieto's lists. The final corpus is thus based on a total of 466 ideophones.

#### **4. Phonological aspects**

This section describes the most salient phonological properties of Sena ideophones. It first delves into the segmental inventory (§4.1), then examines the different syllabic shapes (§4.2), and ends with the suprasegmental level (§4.3).

##### ***4.1 Segmental inventory***

An often-stated criterion in defining the ideophone class is their ability to incorporate phonemes not otherwise attested in the phonological inventory of the language (see, e.g. Samarin 1971: 135–136; Childs 1994; Voeltz & Kilian-Hatz 2001; Dingemanse 2012: 656; Andrason 2017: 146; Andrason 2020: 125–126). Ideophones in Sena do not make use of special phonemes. The segmental inventory in Sena consists of five vowels and 44 consonants. Vowel length is not a contrastive feature in the language, but vowel lengthening plays a crucial role in ideophones, as discussed in §4.2. Each of the five vowels participates in the expression of ideophones. There is no strict rule as to the nature of co-occurring vowels in ideophones. Some ideophones display

identical vowels, as in (3). Some combine high/close vowels (4a), front vowels (4b) or back vowels (4c). Others put vowels with different features together (5).

(3)	i-i	<i>gwíbídí</i>	‘being entered, introduced’
	e-e	<i>ndéngere</i>	‘left aside, useless’
	a-a	<i>dára</i>	‘aged, from a previous round’
	o-o	<i>toro</i>	‘swelling of pregnant belly’
	u-u	<i>gugú:du</i>	‘be dry (soil, spread maize, cloth, heart, etc.)’
(4)	a. i-u	<i>cíŋgu</i>	‘turning around to look behind’
	u-i	<i>cúbwi</i>	‘splashing sound’
	b. e-i	<i>yetí</i>	‘shining, sparkling’
		<i>cipézi-pézi</i>	‘naked’
	c. o-u	<i>gómú</i>	‘negate completely’
		<i>ó:bvu</i>	‘rotten’
(5)	i-o	<i>zikó:</i>	‘very deep’
	o-i	<i>otsí</i>	‘sound of sneezing’
	u-e	<i>búré:</i>	‘straight (for hair)’
	a-u	<i>katú-katu</i>	‘lukewarm (liquid, food)’

Sena has a rich consonantal system including implosives and a large series of prenasalized clusters and affricates. Although some consonants are less frequent than others, as is the case with the other word classes, the whole consonantal inventory is represented in ideophones. Voiced bilabial and dentals are by default implosive. Very few words in the language have the plosives /b/ and /d/. Ideophones are no exception: only four ideophones in the corpus, listed in (6), have them. The minimal (or near) pairs show the contrast between plosives and implosives.

(6)	a. <i>dúdu:ru</i>	‘sound of impact on the ground’
	<i>dudú:ru</i>	‘something turned dry (lake, pan, bone)’
	b. <i>bwí-bwi-bwí</i>	‘making bubbles on the water surface’
	<i>bwí-bwí-bwí</i>	‘throwing food in the mouth’
	c. <i>dí-dí-dí</i>	‘sound of running strides or door knocking’
	<i>dé-dé-dé</i>	‘sleeping on the back’

d. <i>bó(ɔ)</i>	‘spreading out, spilling’
<i>bó:ɔ</i>	‘be damaged’

There does not seem to be any strong correlation between Sena ideophones and sound symbolism (§7.2).

#### 4.2 Syllabic shapes

The different shapes of Sena ideophones are listed in Table 3. The vast majority are not distinct in shape from non-ideophonic words, in that they maintain the canonical CV syllable structure of the language (but see below for exceptions). There are both mono- and polysyllabic ideophones.

Shape	Ideophone	Meaning
monosyllabic	<i>zwí</i>	‘act of throwing’
disyllabic	<i>ná.wa</i>	‘be in a seated position’
	<i>ru.bu.dú</i>	‘with big belly button’
polysyllabic	<i>bo.ndo.kó.te</i>	‘woman sitting with knees bent’
	<i>dzí.ndzí.rí.ki.ti</i>	‘recalling sth. forgotten (& going back)’

**Table 3:** Syllabic structure of Sena ideophones.

Monosyllabic ideophones are proportionally few (81 out of 466, i.e. 17%), but still far more frequent than in other word classes. Most ideophones in Sena are disyllabic. By default, they have a C-initial structure. Vowel-initial ideophones are rare. The whole corpus-based list is provided in Table 4.

Ideophone	Meaning
<i>enwá-enwá</i>	‘being visible (game, fugitive)’
<i>ófu-ófu</i>	‘sound of pig’s grunt’
<i>ó:bvú</i>	‘being rotten’
<i>o:tsí</i>	‘sound of sneezing’
<i>oyó-oyo</i>	‘speaking at the same time’
<i>uyá-uyá</i>	‘movement triggered by strong winds’

**Table 4:** V-initial ideophones in Sena.

Although ideophones overall follow regular phonotactic rules, they may violate certain structure conditions. A first phonological anomaly has to do with the extensive use of vowel lengthening, a feature shared with interjections such as *wá:* (‘exclamation of surprise’), *ʒí:* (‘exclamation of surprise’), *á:y* (‘exclamation of pain’).<sup>5</sup> In a system where long vowels are non-phonemic, vowel lengthening is used for expressive purposes in ideophonic words. With the exception of fully-reduplicated ideophones used to express recurrent, repetitive or rhythmic events (see §7.2), it seems that all ideophones can be subjected to vowel lengthening. Depending on intensity, vowels can be extra-long, with no apparent limit on the length. Lengthening occurs in mono- as well as polysyllabic ideophones as shown in Table 5. Where lengthening occurs in polysyllabic ideophones is not always predictable. In disyllabic ideophones, it may be on the last vowel (e.g. *bídó:*) or the first one (e.g. *dé:fu*). In longer ideophones, lengthening may not apply to the first vowel; it is restricted to either the last (e.g. *bveséré:* and *dodorído:*) or penultimate vowel (e.g. *ndungú:ndu* and *bondokó:te*).

Shape	Ideophone	Meaning
monosyllabic	<i>gwí</i> / <i>gwí:</i>	‘with strength’
disyllabic	<i>bí.dó</i> / <i>bí.dó:</i>	‘dark, black’
	<i>dé.fu</i> / <i>dé:fu</i>	‘be empty (stomach, food bag)’
trisyllabic	<i>bve.sé.ré</i> / <i>bve.sé.ré:</i>	‘listening’
	<i>ndu.ngú.ndu</i> / <i>ndu.ngú:ndu</i>	‘getting calm (body pain, sea wave)’
four-syllable	<i>dó.dó.rí.dó</i> / <i>dó.dó.rí.dó:</i>	‘waiting for s.o. in vain’
	<i>bo.ndo.kó.te</i> / <i>bo.ndo.kó:te</i>	‘woman sitting with knees bent’

Table 5: Vowel lengthening in Sena ideophones.

Examples (7) and (8), extracted from the same conversation, illustrate the use of the ideophone *p<sup>h</sup>wí:* in similar contexts and with equivalent meaning, with and without vowel lengthening.

- (7) *ufa basi udama(ra) kwenda uku, basiya p<sup>h</sup>wí:!*  
*u-fa basi u-da-mara ku-enda uku basiya p<sup>h</sup>wí:*  
 14-flour only AGRP14-PFV.REL-finish INF-go 17.DEM.I 5.basin IDEO(full)  
 ‘(...) flour only which ended up going there, a full basin!’ (Conv.06\_MaMa-BeCh-CeTh#262)

<sup>5</sup> Vowel lengthening in other parts of speech may occur for emphasis or insistence, but it seems much less frequent.

- (8) *ora zonsene unḡafika kwenda kafunḡura basiya unagumana basiya p<sup>h</sup>wi!*  
*ora zi-onsene u-ḡga-fika ku-enda ka-funḡura basiya*  
 10.time AGRP10-all SP2SG-SIT-arrive INF-go AM.ITV-open 5.basin  
*u-na-gumana basiya p<sup>h</sup>wí*  
 SP1-PRS-meet 5.basin IDEO(full)  
 ‘Any time when you arrive and open the basin, you find the basin completely full.’ (Conv.06\_MaMa-BeCh-CeTh\_#268)

In some cases, vowel length participates in the distinction between ideophones. This is interesting because this is the only domain of the language where vowel length plays a distinctive role. Examples of minimal pairs based on vowel length are provided in (9). Again, the length depends on the intensity conveyed.

- (9) a. *dára* ‘aged, from a previous round’  
*dá::ra* ‘in large number’  
 b. *gá* ‘sound of cutting with strength’  
*gá::* ‘rising, ascending (sun, light)’  
 c. *bvu-bvú-bvu-bvu* ‘spending helplessly’  
*bvú::bvú::bvú::* ‘movement of wind or fan’  
 d. *pfé-pfé-pfé* ‘installing/hanging on walls’  
*pfé::* ‘replying unitedly to a song’

Another exceptional feature of Sena ideophones is their ability to end in consonants. This is found with six items only (1,3%), listed in Table 6. Except for *p<sup>h</sup>urú::*, all are monosyllabic. The single occurrence with a nasal coda, *mvú::n*, has a long vowel. All the others end in a long trill [r::] following a short vowel. Free variation between C-final *cú::* and V-final *cú:* suggests that lengthening may occur either on the trill, or on the vowel, but not both. Consonantal codas as well as trill lengthening are not found in other parts of speech.

Ideophone	Meaning
<i>mvú::n</i>	‘smelling bad’
<i>gwú::</i>	‘in abundance’
<i>tír::</i>	‘being squeezed and plentiful’
<i>dú::</i>	i) ‘silent with fear (in front of chiefs)’ - ii) ‘fading away (for plant)’
<i>cú:: ~ cú:</i>	‘spilling (for liquid)’
<i>p<sup>h</sup>urú::</i>	i) ‘unrolling mat’ - ii) ‘trying to fly away (trapped bird)’

Table 6: C-final ideophones in Sena.

Two sentences with C-final ideophones are provided in (10) and (11).

(10) *nyama zidagureye zikanuŋk<sup>h</sup>a, ziri mvú::n!*

<i>nyama</i>	<i>zi-dá-gura = eye</i>	<i>zi-ka-nuŋk<sup>h</sup>a</i>	<i>zi-ri</i>
10.meat	AGRP10-PFV.REL-buy = PRO1	SP10-PST.IPFV-smell.bad	SP10-be
<b><i>mvú::n</i></b>			

IDEO

‘The meat s/he bought was smelling bad, it stinks!’ (Elicitation)

(11) *tsaŋga iri dúr:: na dzuwa*

<i>tsaŋga</i>	<i>i-ri</i>	<b><i>dúr::</i></b>	<i>na</i>	<i>dzuwa</i>
5.weed	SP5-be	IDEO	by	5.sun

‘Weeds faded away because of the sun.’ (Elicitation)

### 4.3 Suprasegmental level

As mentioned in §2, Sena does not have tones, but rather stress or accent associated to the penultimate syllable of phrasal domains, and realized through phonetic properties such as loudness and higher pitch. Amidst this predictable stress system, it is interesting to observe that ideophones have a special prosody which makes the distinction between high tones (H), indicated with an acute accent, and zero tones (Ø). The coexistence of stress-like and tone-like properties within the prosodic system of a language is not unique to Sena. It is discussed in detail by Downing (2019) for Tumbuka, a Malawian Bantu language in which tonal contrasts emerge in ideophones only.

In Sena, many tonal patterns on ideophones can be isolated, as listed and illustrated in Table 7 for non-reduplicated ideophones, and Table 8 for reduplicated ideophones. The tone position in an ideophone does not depend on phonological rules; it is largely unpredictable and must be learnt for each ideophone. The two lists below do not claim to be exhaustive; more research is needed to determine to which extent these tone patterns are strictly assigned to ideophones, and whether they are consistently used across speakers. Crucially, none of these patterns are attested elsewhere, except H:: for interjections. Unlike polysyllabic ideophones, monosyllabic ideophones all seem to

attract a high tone. As can be seen, high-toned syllables may be pronounced longer, but this is not systematic.

Tone pattern	Ideophone	Meaning
1-syllable		
H::	<i>bzwé::</i>	‘dry a little, in part’
HH::	<i>dú::</i>	i) ‘silent with fear (in front of chiefs)’ ii) ‘fading away (for plant)’
2-syllable		
∅.∅	<i>to.ro</i>	‘swelling of pregnant belly’
∅.H	<i>bo.fó</i>	‘be damaged, deformed (following some pressure)’
∅.H::	<i>zi.kó::</i>	‘very deep’
∅.:H	<i>bvu.mbú</i>	‘leaving suddenly’
H.∅	<i>cé.tu</i>	‘glowing, sparkling’
H.:∅	<i>bé.p<sup>h</sup>e</i>	‘experiencing emotional distress’
H.∅::	<i>cé.te::</i>	‘in silence’
H.H	<i>fú.mbú</i>	‘in vain’
H.H:	<i>bí.dó:</i>	‘dark, black’
3-syllable		
∅.∅.∅	<i>kwi.kwi.nya</i>	‘buttocks swaying’
∅.H.H::	<i>bve.sé.ré::</i>	‘listening’
∅.H.H	<i>dzo.ró.ró</i>	‘standing on feet, be vertical, erected’
∅.H.:H	<i>so.ró:.ré</i>	‘tearing, extracting sth. stuck’
H.H::∅	<i>bzwé.ré::.re</i>	‘being smooth (skin, bark, surface)’
H.H.H	<i>jé.ré.ré</i>	‘well lit, burning (light, fire)’
H.∅::∅	<i>dó.ndo::.ro</i>	‘well straight on the way’
4-syllable		
∅.∅.∅.∅	<i>ta.mba.rá:.re</i>	‘sitting with stretched legs’
∅.∅.∅.∅:	<i>do.do.rí.do:</i>	‘waiting for s.o. in vain’
∅.∅.∅.∅	<i>pa.pa.rá.rá</i>	‘extending (wings, arms) while walking’
H.H.H.H	<i>p<sup>h</sup>á.ná.má.ná</i>	‘observing, watching someone’
5-syllable		
H-H-H-∅-∅	<i>dzí.ndzí.rí.ki.t<sup>h</sup>i</i>	‘recalling sth. forgotten (& going back)’

Table 7: Tone patterns in non-reduplicated Sena ideophones.

Tone pattern	Ideophone	Meaning
monosyllabic reduplication		
H-H-H	<i>jó-jó-jó</i>	‘dripping’
H-Ø-H	<i>rú-ru-rú</i>	‘comforting gesture to calm a crying child’
H-Ø-Ø	<i>fá-fa-fa</i>	‘filling up too much, exceeding’
disyllabic reduplication		
Ø.Ø.Ø.Ø	<i>tsa.ri-tsa.ri</i>	‘having last convulsions before dying’
Ø.H-Ø.Ø	<i>ra.sá-ra.sa</i>	‘throwing’
Ø.Ø-H.H	<i>ri.kwi-rí.kwí</i>	‘stuck in the throat’
Ø.H-Ø.H	<i>ca.rí-ca.rí</i>	‘bypassing, deviating’
H.H-H.H	<i>bú.rú-bú.rú</i>	‘in movement (water, liquid)’
H.H-Ø.Ø	<i>fi.dá-fi.dá</i>	‘with lots of cash’
H.Ø-H.Ø	<i>ó.fu-ó.fu</i>	‘sound of pig’s grunt’
trisyllabic reduplication		
Ø.H.H-Ø.H.H	<i>ba.rá.rá-ba.rá.rá</i>	‘walk in an uncontrolled way’
Ø.Ø.Ø-H.Ø.Ø	<i>ke.ŋke.re-ké.ŋke.re</i>	‘sound of foreign language’
Ø.Ø.Ø-Ø.H.H	<i>dzo.ro.ro-dzo.ró.ró</i>	‘standing on feet, be vertical, erected’
Ø.Ø.H-Ø.Ø.Ø	<i>pe.ke.té-pe.ke.te</i>	‘movement of confusion, disorder (in the forest)’

Table 8: Tone patterns in reduplicated Sena ideophones.

Ideophonic tones can be contrastive and contribute to semantic variation. For instance, the three ideophones in (12) are all derived from the verb stem *beruka* ‘surface, float’ (see §5.1 for deverbal ideophone formation), but differ in their syllabic shapes as well as their tone pattern. Very often, vowel lengthening is combined with a H tone on the syllable (but cases of toneless long vowels such as *bvu:mbú* ‘leaving suddenly’, *céte::* ‘in silence’ or *dóndo::ro* ‘well straight on the way’ in Table 7 also exist). This leads to two tone patterns in case of disyllabic ideophones: H::Ø as in *bé::ru* or Ø.H:: as in *berú::*. Reduplicated (or triplicated) patterns usually do not involve vowel lengthening, and the assignment of the tone-bearing syllable is less predictable. The tone pattern Ø.H-Ø.H-Ø.H is found in *berú-berú-berú*.

- (12) *bé::ru* ‘emerging, surfacing’ = from bottom to the surface  
*berú::* ‘floating on the surface’ = at the surface already  
*berú-berú-berú* ‘moving back and forth on the surface’ = with movements  
 < verb *beruka* ‘surface, float’



As can be seen, the meaning of these three ideophones is very close. It seems that the H::Ø tone pattern (*bé::ru*) depicts the length of the action of floating up, until it eventually reaches the end point (the surface). In contrast, the Ø.H:: pattern (*berú::*) depicts an unbounded action. The regular Ø.H-Ø.H-Ø.H pattern in triplicated *berú-berú-berú* seems to convey the rhythm of movement. Further research would be needed to check whether there is any correlation between the tone pattern and the lexical aspect the ideophone encodes. Tone variation alone is, however, not responsible for the nuances of interpretation observed in (12). Nor is it associated to specific meaning, as reported, e.g. for Ewe (ewe), where raised pitch (high tone) is symbolic of niceness, pleasantness, goodness and small dimension, whereas unraised pitch (low tone) indicates badness, unpleasantness and big dimension (Ameka 2001: 30).

## 5. Morphological aspects

### 5.1 Derivation

Compared to other word classes, Sena ideophones display very little morphology, a characteristic shared by most African languages with ideophones (Childs 1994: 185). First, Sena ideophones are uninflected words. Unlike adnominal modifiers and (most) verb forms, they do not take agreement prefixes. This suggests a stronger syntactic independence, or “syntactic aloofness” in Kunene’s (1978: 13) terms.

As for derivational morphology, two types of ideophones may be distinguished: (i) those which have the shape of primary roots with no derivational affixes (e.g. *rámhá* ‘lying, resting, sleeping’ or *zýú::* ‘without return, for good’) and (ii) those which are derived from verbs. The rest of this section is devoted to the second type.

The majority of ideophones in Sena have the shape of verb roots, i.e. verbs without derivational and inflectional suffixes (including the final vowel *-a* of the verb, the applicative *-ir* ~ *-er*, the causative *-is* ~ *-es*, the neuter/reflexive *-ik* ~ *-ek*, and the reciprocal/associative *-an*). Two main patterns emerge. In the first, ideophones copy full verb roots as in (13) for simple ideophones and (14) for reduplicated ideophones. Since Sena phonotactics forbids syllable codas (with the exception of a handful of ideophones listed in Table 6, §4.2), a final vowel is added to C-final verb roots. The value of this added vowel is difficult to predict. If disyllabic (or more) verb roots have identical vowels, the added vowel often harmonizes (e.g. in *t<sup>h</sup>índini*, *t<sup>h</sup>úkutu* and *barára-barára*), but this rule of vowel harmonization does not always hold, as e.g. in *funukúré*. The choice for the vowel added to ideophones derived from monosyllabic verb roots (*zikó::*, *redé-redé*, *bwekú-bweku*, *topó-topo*) does not seem to follow any principle.

## (13) Simple full verb root

<b>zikó::</b>	‘very deep’
< <i>ku-zik-a</i>	‘to be deep’
<b>funukúré</b>	‘movement of opening’
< <i>ku-funukur-a</i>	‘to open’
<b>t<sup>h</sup>índini</b>	‘in silence, refusing to talk’
< <i>ku-t<sup>h</sup>índin-a</i>	‘to be silent, refuse to talk’
<b>t<sup>h</sup>úkutu</b>	‘landing carefully’
< <i>ku-t<sup>h</sup>ukutiz-a</i>	‘to land carefully’

## (14) Reduplicated full verb root

<b>barára-barára</b>	‘walk in an uncontrolled way’
< <i>ku-barar-ik-a</i>	‘to be uncontrolled’
<b>redé-rede</b>	‘as equals’
< <i>ku-red-an-a</i>	‘to do sth. in the same way’
<b>bwékú-bweku</b>	‘speaking a lot and unnecessarily’
< <i>ku-bwek-a</i>	‘to speak a lot and unnecessarily’
<b>topó-topo</b>	‘falling asleep’
< <i>ku-top-a</i>	‘to get very tired, fall asleep’

In the second pattern, illustrated in (15)-(16), ideophones copy only the first part of verb roots but leave out the coda. This corresponds to the initial CV segments of monosyllabic roots or the initial CVCV segments of disyllabic (or trisyllabic) roots. Again, reduplication (or triplication) may operate (16) or not (15). Vowel lengthening is more prone to occur in non-reduplicated ideophones (see §7.2).

## (15) Simple partial verb root

<b>bofó</b>	‘be damaged, deformed (following some pressure)’
< <i>ku-bofor-a</i>	‘to damage, deform, squeeze (following some pressure)’
<b>bvú::</b>	‘heating, warming up (sun)’
< <i>ku-bvuk-a</i>	‘to heat, warm up (sun)’
<b>bzwé::</b>	‘dry a little, in part’
< <i>ku-bzwer-er-a</i>	‘to dry a little, in part’

## (16) Reduplicated partial verb root

<b>jó-jó-jó</b>	‘dripping’
< <i>ku-joj-a</i>	‘to drip’

<b>derú-derú</b>	‘swinging, rocking’
< <i>ku-derump</i> <sup>h-a</sup>	‘to swing, rock’
<b>wá-wá-wá</b>	‘pouring and spreading (salt in pan, water on land)’
< <i>ku-wadz-a</i>	‘to pour and spread (salt in pan, water on land portion)’

Following Steriade’s (1988) formal account of reduplication, referred to in Ameka (2001: 31), it is likely that the second pattern above (partial reduplication) derives from the first pattern (full reduplication), i.e. reduplication of verb roots is total, and then followed by pruning of the final segment of the reduplicated material.

More rarely, some segmental material is added to the derived ideophone, as in (17), but the examples are too few to make any generalizations.

(17) <b>dyókó-dyoko</b>	‘craving for food’
< <i>ku-dy-a</i>	‘to eat’
<b>timbwi-timbwi</b>	‘with tightened belt; suffering because of s.o.’
< <i>ku-timb-a</i>	‘to tighten too much (rope, belt)’

Among these two examples, the shape of *timbwi-timbwi* is intriguing, as it looks as if the passive suffix *-(i)w* was added to the verb root *timb*. Similarly, the recurring segment *erere* in the four ideophones *t<sup>h</sup>emberé::re* ‘sick in bed’, *remberé::re* ‘about to fall’, *jérére* ‘being alight’ and *bzwéré::re* ‘being smooth (skin, bark, surface)’, suggests the presence of the applicative morpheme *-er*. Unlike *timbwi-timbwi*, synchronic underived verb forms are not attested for these four ideophones. Even though it is not unusual to have verbs with applicative morphology but no verbal base, a verbal origin might still be postulated. The possible presence of passive and applicative morphology in a handful of ideophones is interesting in two respects. First, it would provide some exception to the generalization that ideophones have no derivational morphology. Second, it would confirm that this type of ideophones is indeed derived from verbs, and not the other way around. This is in line with what a number of Africanist scholars have proposed for other languages, including Meyer (this volume), Courtenay (1976), Childs (1994) and Doke (1954), although the latter also advocate for a two-way derivation (verb > ideophone and ideophone > verb) in Lamba (lam).

Deverbal ideophones are often used with the verb to which they are morphologically related, as in (18), where the ideophone *bwandzú* ‘being open, uncovered, discovered’ follows the conjugated verb *bwandzuka* ‘be open, uncovered,

discovered’ (see §6.1 for more details on this construction). This is, however, not systematic. In (19), the same ideophone is used without its cognate verb.

- (18) *ŋkazi ure, ŋguwo zace zaɓwandzuka na mp<sup>h</sup>epo ɓwandzú*  
*ŋ-kazi ure ŋguwo zace za-ɓwandzuka na mp<sup>h</sup>epo*  
 1-woman 9.DEM.III 9.cloth 9.POSS3SG SP9.PFV-be.uncovered with 9.wind  
***ɓwandzú***

IDEO(being uncovered)

‘That woman, her cloth has been uncovered by the wind, [IDEO].’ (Elicitation)

- (19) *ŋsuwo uri ɓwandzú*  
*ŋ-suwo u-ri ɓwandzú*  
 3-door SP3-be IDEO(being open)  
 ‘The door is open.’ (Elicitation)

The derivation of ideophones from verbs is very productive in Sena, and it is not clear whether there are any restrictions in terms of semantics or types of verbs. From the available data, it rather looks as if almost any verb could yield an ideophone with an associated depictive function. Possible exceptions exist with verbs which already trigger the formation of adverb-like words in *ci-* (§2). The two forms in (20) formally differ in the presence or absence of the derivational prefix *ci-*. Whereas the adverb *cikwatakwata* morphologically and semantically derives from the verb stem *kwata* ‘hold, take’, the origin of the ideophone *kwatakwata* ‘leaning’ and its semantic assignment are unknown. It is not possible to predict at this stage when verbs will be derived adverbially with *ci-* and when they use the “bare” ideophonic verb form only.

- (20) *ci-kwata-kwata* ‘willing to touch and take everything around’  
*kwata-kwata* ‘leaning’

It is also not clear from the corpus whether verbs or nouns can be derived from ideophones. The few verbs listed in (21) are possible candidates. Notably, the ideophones they (possibly) derive from are all onomatopoeic. This is particularly evident for animal sounds, such as *wú-wu-wú* and *mé-mé*. In these two cases, a deverbal derivational direction seems less likely. The direction of the last three pairs of examples is less obvious, but their mimetics is such that it is easy to conceive a de-ideophonic derivational process. Onomatopoeic ideophones appear as relevant sources to create new verbs.

- (21) *ku-wut-a* ‘to bark’  
 < *wú-wu-wú* ‘sound of dog barking’  
*ku-memes-a* ‘to bleat’  
 < *mé:-mé:* ‘sound of goat bleating’  
*ku-fot-a* ‘to make a silent fart’  
 < *fó:* ‘sound of a silent fart’  
*ku-gagaḍ-a* ‘to cut in pieces with a hoe, machete, knife’  
 < *gá* ‘sound of cutting with strength, in one go’  
*ku-nyenyen-a* ‘to gnaw’  
 < *nyenyena* ‘gnawing’

## 5.2 Reduplication (and multiplication)

Many ideophones are formed by total reduplication (*gabi-gabi*), which may extend to triplication (*fyé-fye-fye*) and even multiplication (*bvu-bvú-bvu-bvu*). Albeit rarer, partial reduplication is also attested, either initial reduplication (*kwikwinya*) or final reduplication (*dzoróró*). Submorphemic segments such as *erere* (with or without vowel lengthening), occurring in *t<sup>h</sup>emberé::re*, *remberé::re*, *jérére* and *bzwéré::re*, are rare.

Shape	Ideophone	Meaning
total reduplication	<i>rwá-rwá</i>	‘with eyes wide open, without blinking’
	<i>gabi-gabi</i>	‘movement of choppy water’
	<i>p<sup>h</sup>eperú-p<sup>h</sup>eperu</i>	‘staggering’
triplication	<i>fyé-fye-fye</i>	‘staying mute, not reporting a culprit’
multiplication	<i>bvu-bvú-bvu-bvu</i>	‘spending helplessly’
first syllable repeated	<i>kwikwinya</i>	‘buttocks swaying’
last syllable repeated	<i>dzoróró</i>	‘standing on feet, be vertical, erected’
with recurring segments	<i>t<sup>h</sup>emberé::re</i>	‘sick in bed’
	<i>remberé::re</i>	‘about to fall’
	<i>jérére</i>	‘being alight’
	<i>bzwéré::re</i>	‘being smooth (skin, bark, surface)’

Table 9: Reduplication patterns in Sena ideophones.

Reduplication is a recurrent derivational mechanism in Sena. As can be seen from Table 9, reduplication may be full or partial. Reduplication of reduplication, or double reduplication, resulting in four identical segments, is also commonly attested. Full

reduplication involves base forms which have the segmental structures CV (*rwá-rwá* ‘with eyes wide open, without blinking’), VCV (*oyó-oyo* ‘speaking at the same time’), CVCV (*gabí-gabí* ‘movement of choppy water’) or CVCVCV (*p<sup>h</sup>eperú-p<sup>h</sup>eperu* ‘staggering’). While reduplication applies to disyllabic (or trisyllabic) ideophones, triplication most often targets monosyllabic ideophones. Since Sena has both prefixal and suffixal morphology, it is difficult to determine the base and the reduplicated part. Reduplication of the otherwise very rare CVC pattern (§4.2) is not attested. Partial reduplication of ideophones involves reduplication of the final CV (*dzoróró*) or the initial CV of the base form (*kwikwinya*).

Total reduplication is not the exclusive property of ideophones in Sena. It is a recurrent morphological process also attested in nouns (22), adjectives (46), verbs (23) and adverbs (24). However, only ideophones allow partial reduplication, triplication and multiplication.

(22) *kunacitao ṛsonk<sup>h</sup>o, kunaburuka ṛsonk<sup>h</sup>o-ṛsonk<sup>h</sup>o, ndi kweko kunap<sup>h</sup>atene basa.*

*ku-na-cita = o                      ṛsonk<sup>h</sup>o    ku-na-buruka                      ṛsonk<sup>h</sup>o-ṛsonk<sup>h</sup>o*

AGRP17-PRS-do.REL = PRO2    10.tax    AGRP17-PRS-go.out.REL    10.tax-RED

*ndi kweko    ku-na-p<sup>h</sup>at = ene                      basa*

COP    17.DEM.II    AGRP17-PRS-take.REL = PRO1SG    5.work

‘Where they charge taxes, where taxes are levied (go out), it is where I work.’

(Life.Story\_DoDa\_#89)

(23) *onoyu anak<sup>h</sup>ara bwera kandimenya-menya pano*

*onoyu    a-na-k<sup>h</sup>ara                      bwera    ka-ndi-menya-menya    pano*

1.DEM.III    AGRP1-PRS-be.REL    come    AM.ITV-OP1SG-beat-RED    16.DEM.I

‘That one who keeps coming and beating me here.’ (Story\_HareTurtle\_#88)

(24) a. *cinjala-njala*                      ‘always with hunger’

< *njala*                      ‘hunger’

*cimanja-manja*                      ‘with empty hands’

< *manja*                      ‘hands’

*cip<sup>h</sup>ata-p<sup>h</sup>ata*                      ‘willing to touch everything around’

< *kup<sup>h</sup>ata*                      ‘to take, seize’

*cisuzi-suzi*                      ‘looking with curiosity’

< *kusuzumira*                      ‘to observe, watch, spy on’

## 6. Syntactic and distributional aspects

This section explores the syntactic and distributional aspects of ideophones and the different degrees of syntactic integration they exhibit. Before doing so, it is worth mentioning that ideophones in Sena are predominantly used in affirmative declarative utterances, but also frequent in subordinate clauses, especially relatives. This is not the case with other utterance types such as questions, imperatives and negatives, where ideophones are not attested, except for the negative sentence in (25).

- (25) *ŋk<sup>h</sup>abe kwenda kwene gumu-gumu tayu, kwenda kabvundzira*  
*ŋk<sup>h</sup>abe ku-enda ku-ene gumu-gumu tayu*  
 NEG INF-go AGRP15-INT IDEO(walking insecurely palpating) NEG  
*ku-enda ka-bvundzira*  
 INF-go AM.ITV-ask.for  
 ‘(...) to go dating (lit. asking for) is not about going randomly here and there (lit. it is not going randomly here and there, go dating).’ (Poem\_Kubvundzira)

### 6.1 Collocational ideophones

Collocational ideophones are those which co-express the event expressed by a verb. They function similarly to manner adverbs or converbs (defined as dependent verbs marking adverbial subordination). Most often, the ideophone is deverbal and combines with its source verb, as in (26)-(27), where *bó::* and *cé::* follow a conjugated form of the verb roots *bom* ‘spill’ and *cek* ‘cut’, respectively. In these constructions, ideophones are to a certain extent similar to so-called “cognate objects” (Jones 1988; Levin 1993), with the difference that they do not exhibit nominal features. Since they denote the same meaning as the verb, they are also optional. Their function is merely expressive.

- (26) *madzi abomeka bó::!*  
*ma-dzi a-bom-ek-a bó::*  
 6-water SP6.PFV-spill-NTR-FV IDEO(spilling)  
 ‘The water spilled/spread.’ (Elicitation)

- (27) *iyé akaceka cé:: mbaramba*  
*iyé a-ka-ceka cé:: mba-ramba*  
 PRO3SG SP1-PST.IPFV-cut IDEO(act of cutting) CVB.SP2-fall  
 ‘He was cutting, and they fell down.’ (Story\_HareLion\_#10)

By default, the ideophone immediately follows its cognate verb, even if the latter has a complement. In (28), *nyenyena* intervenes between the verb and its object *pyat<sup>h</sup>u* ‘our belongings’. The privileged after-verb position of the ideophone is, however, not mandatory. In (29), the ideophone *ɓwandzú* appears sentence-finally and is separated from its cognate verb by the complement *na mp<sup>h</sup>epo* ‘by the wind’.

- (28) *tiri kupinyenyena nyenyena pyat<sup>h</sup>umbo tepo kupidyá*  
*ti-ri ku-pi-nyenyena nyenyena pi-at<sup>h</sup>u = mbo*  
 SP1PL-be INF-OP8-gnaw IDEO(gnawing) 8-belonging(POSS1PL) = too  
*tepo ku-pi-dya*  
 like.this INF-OP8-eat  
 ‘We are gnawing our belonging too, eating it this way.’  
 (Conv.05\_BeCh&CeTh\_#24)

- (29) *ɲkazi ure, ɲguwo zace zaɓwandzuka na mp<sup>h</sup>epo ɓwandzú*  
*ɲ-kazi ure ɲguwo zace za-ɓwandzuka na mp<sup>h</sup>epo*  
 1-woman 9.DEM.III 9.cloth 9.POSS3SG SP9.PFV-be.uncovered by 9.wind  
***ɓwandzú***  
 IDEO(being uncovered)  
 ‘That woman, her cloth has been uncovered by the wind.’ (Elicitation)

Collocational ideophones may also combine with non-cognate verbs. These verbs can be semantically close, but not morphologically related. In (30), the ideophone *dwá::ra*, used to characterize something dried, combines with the verb *wuma* ‘dry’. Again, the ideophone is optional, and again, it is located between the verb and its complement.

- (30) *mapira awuma dwá::ra na dzuwa*  
*ma-pira a-wuma dwá::ra na dzuwa*  
 6-maize SP6.PFV-dry IDEO(very dry) with 5.sun  
 ‘Maize dried with the sun.’ (Elicitation)

More exceptionally, ideophones may precede the verbs to which they are semantically related, as in (31), with *káwo* preceding *at<sup>h</sup>awa suro*. Crucially, in this example the prototypical SV structure is also reversed to VS, literally ‘fleeing fled the hare’. It is likely that this permuted word order is pragmatically determined (further research is needed on this topic).



- (31) *pidamara kudya iye nyimo zonsene gédé | káwo at<sup>h</sup>awa suro*  
*[pi-da-mara ku-dya iye nyimo zi-onsene gédé]<sub>SUB</sub>*  
 AGRP8-PFV.REL-finish INF-eat PRO3SG 10.bean AGRP10-all IDEO(completed)  
*[káwo a-t<sup>h</sup>awa suro]<sub>MAIN</sub>*  
 IDEO(fleeing) SP1.PFV-flee 1a.hare  
 ‘After having eaten all the beans, the hare fled.’ (Story\_HareTurtle\_#58)

Finally, manner-adverbial ideophones may modify verbs which are not semantically linked. This is the case in the verses of a poem in (32), where the verb form *tinafika* ‘we (will/are to) arrive’ governs the preposed ideophone *nsanga-nsanga* as a manner adverb.

- (32) *t<sup>h</sup>angwi ndife ana ako*  
*tinadzabwerera*  
*nsanga-nsanga tinafika mbatik<sup>h</sup>ariratu*  
*t<sup>h</sup>angwi ndife ana ako ti-na-dza-bwer-er-a*  
 because COP.PRO1PL 2.child AGRP2.POSS2SG SP1PL-PRS-AM.VTV-return-APPL-FV  
*nsanga-nsanga ti-na-fika mba-ti-k<sup>h</sup>ar-ir-a = tu*  
 IDEO(quickly)-RED SP1PL-PRS-arrive CVB-SP1PL-stay-APPL-FV = EMPH  
 ‘because we are your children | we are to come back | as soon as possible we are to come and remain for good.’ (Poem\_SenaSena)

## 6.2 Complements of copula verbs or light verbs

Ideophones may appear as complements of copula verbs or light verbs. In such combinations, the meaning and subcategorization properties are no longer conveyed by the verb, but by the ideophone itself. Complex predicates involving a semantically fairly empty verb and an ideophone are cross-linguistically widespread (cf., among others, Childs 1994: 187; Schultze-Berndt 2001: 360–367; Güldemann 2008: 280; Dingemanse 2012).<sup>6</sup>

Sena has two copula verbs: *ri* and *k<sup>h</sup>ara*. The copula verb *ri* is defective, with inflection markers limited to subject information and two TAM distinctions, namely zero-marked present and imperfective past *k<sup>h</sup>a-* ~ *ka-*. The copula *ri* is typically used to express identity and inclusion predication, introducing nominal and adjectival predicates. Ideophones introduced by *ri* are thus used predicatively, a property

<sup>6</sup> Also see the contributions by Meyer, by Treis and by Authier in this special issue, which all discuss these constructions for their languages.

otherwise shared by nouns and adjectives. Many examples of *ri* + ideophone are available in the corpus, see (10), (11) and (19). Another example is provided in (33) with *ri* + *pyá-pya-pya*. Note that the presence of the copula verb is optional. When absent, the ideophone becomes a holophrastic predicate on its own (see §6.3), as is *gúgúdú* ‘being dry’ in the same example.

(33) *agumana ari pyá-pya-pya, ama' uma zá gúgúdú*

*a-gumana a-ri pyá-pya-pya a-mala uma zá*  
 SP1.PFV-meet SP1-be IDEO(wrapping)-RED SP1.PFV-finish dry already

***gúgúdú***

IDEO(being dry)

‘He found (it), it was wrapped up and had already dried.’ (Story\_HareLion\_#45)

The semi-copula *k<sup>h</sup>ara* ‘be, stay’ is used both predicatively and as a copula in non-verbal predication. Unlike *ri*, *k<sup>h</sup>ara* behaves as a regular verb (or almost) in terms of morphology, and replaces *ri* when TAM features other than present and imperfective past are expressed. In contrast to *ri*, examples of *k<sup>h</sup>ara* + ideophone are rare. Only three occurrences were found, one of which is given in (34).

(34) *ap<sup>h</sup>are akamenya bora, pidafika asupayi, onsene ak<sup>h</sup>ara dúr::!*

*a-p<sup>h</sup>are a-ka-menya bora pi-da-fika*  
 2-youth SP2-PST.IPFV-beat 5.ball AGRP8-PFV.REL-arrive

*a-supayi a-onsene a-k<sup>h</sup>ara dúr::!*  
 2-policeman AGRP2-all SP2.PFV-be/stay IDEO(silent in fear)

‘The youth were playing football; when the police arrived, all stayed silent in fear.’  
 (Elicitation)

It also happens that the occurrence of ideophones is conditioned by individual verbs, which function as “dummy” or “introductory” verbs (Childs 1994: 187), best known in typology as light verbs (Jespersen 1965). The basic meaning of these ideophonic complex predicates relies on the ideophone and not on the supporting verb, whose original meaning tends to be bleached, as the English verb ‘take’ in ‘take a walk’. At least two verbs fulfill this role in Sena, namely *cita* ‘do, make’ and *bveka* ‘feel, hear’. While the combination *cita* + ideophone is highly productive, the use of the light verb *bveka* appears to be more restricted. In (35), the only example retrieved from the corpus, *bveka* combines with the ideophone *cúbwi*

‘splash’ to express ‘plunge’ (literally ‘hear splash’).<sup>7</sup> It is likely that other ideophones expressing a sound may be used with the light verb *bveka*. It is also likely that *cita* would be accepted in the same context.

- (35) *iye bveka cúbwi muñcera mure*  
*iye      bveka              cúbwi              mu-ñ-cera      mure*  
 PRO3SG feel/hear.NARR IDEO(splash) 18-3-well 18.DEM.III  
 ‘He plunged into that well.’ (Story\_HareTurtle\_#96)

The light verb *cita*, on the other hand, seems to combine with any ideophone to express actions that are normally denoted by dynamic verbs. In (36), *cita* + *gwibídi* means ‘enter’. It is naturally followed by a locative complement. In these constructions, the light verb *cita* does not serve, however, to predicate qualities or other stative situations. In this sense, one might conclude that *cita* used with ideophones has not completely lost its original semantics and still conveys a dynamic interpretation. These ideophonic constructions thus differ from those based on copula verbs seen above which feature identity and inclusion predication.

- (36) *ñkazi anɔacita gwibídi ñnyumba, menyewatu*  
*ñ-kazi      a-ñga-cita      gwibídi                              ñ-nyumba*  
 1-woman SP1-SIT-do IDEO(being entered, introduced) 18-9.house  
*meny-er-w-a = tu*  
 beat.NARR-APPL-PASS-FV = EMPH  
 ‘The woman, as soon as she gets home, is beaten right away.’ (Conv.06\_MaMa-BeCh-CeTh\_#366)

Complex predicates based on *cita* + ideophone are frequently attested in relatives. Two consecutive relative clauses based on this combination are provided in (37). In the first, *cita* + *zwi* express ‘throw’. In the second, *cita* applies to two subsequent ideophones, *nt<sup>h</sup>úngúru* ‘sound of joy and satisfaction’ and *mezi* ‘way of swallowing any food’. This last case is interesting in that it shows that ideophones can be juxtaposed. Note that no pause or prosodic break between *nt<sup>h</sup>úngúru* and *mezi* is heard, probably because they participate in the same event.

---

<sup>7</sup> Note that bare verb stems, i.e. without subject and TAM marking, as *bveka* in (35), are frequently heard in Sena narratives when a sequence of events occurs (hence they are glossed NARR for narrative). Other examples are found in (41) with *fika* ‘he arrived’ and in (62) with *funa* ‘it wanted’.

(37) *pinaciteye zwí muñcera mure, pinaciteye nt<sup>h</sup>úngúru mezi ...*

*pi-na-cit* = eye                      *zwí*                      *mu-ñ-cera*    *mure*  
 AGRP8-PRS-do.REL = PRO1    IDEO(throwing)    18-3-well    18.DEM.III

*pi-na-cit* = eye                      *nt<sup>h</sup>úngúru*                      *mezi*  
 AGRP8-PRS-do.REL = PRO1    IDEO(sound of joy)    IDEO(swallowing)

‘When he throws (it) in that well, and when he noisily (and) happily swallows (it), ...’ (Story\_HareTurtle\_#134)

Although the combination *cita* + ideophone generally yields intransitive constructions, it may be used transitively. In (38), the noun phrase *suro ure* ‘that hare’ functions as the object licensed by the predicative construction *cita cíngu* ‘turn around to look behind’.

(38) *pidáfika iye añfikira na kuk<sup>h</sup>odoro, pinaciteye cíngu suro ure ...*

*pi-dá-fika*                      *iye*                      *a-ñ-fik-ir-a*                      *na ku-k<sup>h</sup>odoro*  
 AGRP8-PFV.REL-arrive    PRO3SG    SP1-OP1-arrive-APPL-FV    by 17-back

*pi-na-cit* = eye                      *cíngu*                      *suro*    *ure*  
 AGRP8-PRS-do.REL = PRO1    IDEO(turning around to look behind) 1a.hare 1.DEM.

‘When he (the hare) arrived, he arrived behind his back. When he (the turtle) turns around and looks at that hare behind ....’ (Story\_HareTurtle\_#50)

Crucially, in all these examples, the ideophone cannot be separated from the light verb by any other constituents. This rigid order constitutes further evidence that they form a syntactic and semantic unit.

### 6.3 Ideophonic predicates

Very often, ideophones occur in verbless clauses and act as holophrastic predicates by themselves. The event is expressed by the ideophone, which does not take any verbal morphology and functions as the comment of a topic-comment sentence. The ideophone also determines the argument structure of the clause, in that it controls the semantic roles assigned to the noun phrases present in the clause, a function typically fulfilled by verbs.

More often, it combines with a single noun phrase assuming the semantic role of agent. In (39), *míyó:* (< verb stem *miyonga* ‘flee’) and *p<sup>h</sup>é:* designate the act of fleeing and hiding and the act of staying still and watching silently, respectively. The action

is performed by the agentive participant *kamba* ‘turtle’, which functions as the subject argument of both ideophonic predicates.

The whole sentence is verbless: ideophones express events which would otherwise be expressed by inflected verbs in a sequence of independent clauses, and the remaining constituents express the participants of the event (i.e. *iyé kamba* ‘he the turtle’ in (39) as subject argument) and optionally, adjuncts (i.e. *ḡk<sup>h</sup>undu-ḡk<sup>h</sup>undu mwa ḡjira mure* ‘on that side of the path’ as locative adjunct).

- (39) *entāo iyé kamba míyó:: ḡk<sup>h</sup>undu-ḡk<sup>h</sup>undu mwa ḡjira mure p<sup>h</sup>é::*  
*entāo iyé kamba míyó:: ḡk<sup>h</sup>undu-ḡk<sup>h</sup>undu*  
 then PRO3SG 1a.turtle IDEO(fleeing to hide) 18-9.side-RED  
*mu-a ḡjira mure p<sup>h</sup>é::*  
 AGRP18-CONN 9.path 18.DEM.III IDEO(staying still and watch)  
 ‘Then, (he) the turtle fled and hid on that side of the path, he stayed still and watched.’ (Story\_HareTurtle\_#90-91)

Very often, the ideophone forms an independent (verbless) clause on its own, and the previously expressed agent does not need to be repeated. In (40), the ideophone *tsari-tsari* is used alone to depict the action of convulsing before dying endured by the cobra (*nyoka*), mentioned previously in the sentence. In (41), the ideophone *kwí::*, which depicts the action of entering discretely, is integrated into a sequence of events performed by the turtle, whose identity is known from context.

- (40) *ḡkazace ure azakabuka mbasiya nyoka iri pepare tsari-tsari*  
*[ḡ-kazi = ace ure a-za-ka-buka]<sub>MAIN</sub> [mba-siya nyoka]<sub>SUB</sub>*  
 1-woman = POSS3SG 1.DEM.III SP1-AM.VTV?-PST.IPFV-go.out CVB-leave 9.cobra  
*[i-ri pepare]<sub>INDPDT</sub> [tsari-tsari]<sub>INDPDT</sub>*  
 SP9-be 16.DEM.III IDEO(having last convulsions before dying)  
 ‘That woman had just gone out and left the cobra (it is) there having its last convulsions.’ (Story\_AvengingCobra\_#79)

- (41) *fika pat<sup>h</sup>endere kwí:: iyé amona suro akwenda*  
*[fika pa-t<sup>h</sup>endere]<sub>INDPDT</sub> [kwí::]<sub>INDPDT</sub>*  
 arrive.NARR 16-5.bush IDEO(fit in, enter to hide)  
*[iyé a-mu-ona suro]<sub>MAIN</sub> [a-ku-enda]<sub>SUB</sub>*  
 PRO3SG SP1.PFV-OP1-see 9.hare SP1-PRS.PROG-go

‘He arrived at the bush, entered in it not to be seen. He saw the hare going.’  
(Story\_HareTurtle\_#96)

In (42), the ideophone *dedzúdedzú* refers to the act of flapping around when out of one’s environment, typically used for caught fish or birds. In the preceding sentence, we have learned that the hare ate the bait and got caught on the hook. Notably, it is employed with the comitative pronominal phrase *naye*, literally ‘with him’, which is a recurrent device in Sena oral speech, mostly found after verbs (more research is needed to better understand its function).

(42) ... *amaŋga* | *dedzúdedzú naye* | *pinam̩<sup>h</sup>useye kuŋʒa kure*, "*caa, ɓaa kamba ndiwe!*"

*[a-maŋga]*<sub>INDPDT</sub> *[dedzú-dedzú* *na-iye]*<sub>INDPDT</sub>

SP1.PFV-tie IDEO(flapping.around)-RED COM-PRO3SG

[*pi-na-m̩-ɓ<sup>h</sup>usa = iye* *kuŋʒa* *kure*]<sub>SUB</sub>

AGRP8-PRS-OP1-pull.REL = PRO1 17.out 17.DEM.III

*[caa ɓaa kamba ndiwe]*<sub>INDPDT</sub>

INTERJ friend.VOC 9.turtle COP.PRO2SG

‘... he (the hare) got caught. He flapped around desperately. When he (the turtle) pulls him out there: (the hare said) “Oooh, my friend Turtle, it’s you?!”  
(Story\_HareTurtle\_#135)

Note that prosody is not always indicative of the independent status of ideophonic predicates: whereas a neat break (represented by | in the first line) is heard around *dedzúdedzú naye* in (42), *tsari-tsari* in (40) and *kwú:* in (41) are not prosodically detached from the rest of the sentence.

In addition to subject arguments, certain ideophonic predicates may also take objects with the role of theme or instrumental-comitative complements. In (43), the theme participant *kaderace* ‘his chair’ functions as the object of the following ideophonic predicate *kwê:* ‘pushing’, and the transitive construction literally reads as ‘the chair (he) pushed’. Note that depending on their discourse topicality, it is common for objects constituents to occur before the lexical verb. Although a dedicated study on information structure remains to be done, the object - ideophonic predicate order in (43) seems to instantiate a predicate-centered focus. In the same sentence, the second ideophone *náwa* ‘sitting’ also functions predicatively, but only takes the (agentive) subject argument mentioned earlier in the story, i.e. the lion.

- (43) [*panga ŋkazace “ndo’ kap<sup>h</sup>ike ine ndinak<sup>h</sup>ara pano”*,] *kaderace kwê: náwa*  
**kaderace**            **kwê:**            **náwa**  
 9.chair.POSS3SG    IDEO(pushing)    IDEO(sitting)  
 ‘(He (the lion) said to his wife: “go and cook, I’ll stay here!”,) he pushed the chair and sat.’ (Story\_HareLion\_#28)

As in (37), the ideophones in (43) are juxtaposed without a pause between them. In (44), the ideophone *fifu* ‘pulling the hook (when catching fish)’ takes the instrumental adjunct *nao* ‘with it’ referring to the fishing pole mentioned in the preceding clause. Again, the adjunct precedes the ideophone. This order is not rigid though. In (45), for instance, the adjunct *na manyadzo* ‘with shame’ follows the ideophonic predicate *tsuri-tsuru*.

- (44) *iyé akamedza, pidamangeye nyama mbafifura ntete wace, nao fifu!*  
*iyé*    *a-ka-medza*    *pi-da-manga = eye*            *nyama*  
 PRO3SG    SP1-PST.IPFV-fish    AGRP8-PFV.REL-catch = PRO1    9.fish  
*mba-fifura*            *ntete*            *o-ace*            **nao**            **fifu**  
 CVB.SP1-pull.hook    3-fishing.pole    AGRP3-POSS3SG    with.PRO3SG    IDEO(pulling hook)  
 ‘He was fishing. When he caught a fish, he pulled his fishing pole, he pulled the hook with it.’ (Elicitation)

- (45) *ntsikana, tsuri-tsuru na manyadzo, nde munaŋga iye...*  
*nt-sikana*    **tsuri-tsuru**                            **na**    **ma-nyadzo**  
 1-girl    IDEO(looking down shamefully)    with    6-shame  
*nde mu-na-ŋga*            *iyé*  
 COP    AGRP18-PRS-speak.REL    PRO3SG  
 ‘The girl, looking down with shame, it is then she said...’  
 (Poem\_Pakubvundzira)

Ideophonic predicates may be part of complex sentences where they function as the main (verbless) clause embedding a subordinate clause. In (46), the ideophone *odzí-odzí-odzí*, conveying the concept of walking feebly, without strength, functions as the main (verbless) clause introduced by the adverbial subordinate clause of condition/time *angafamba* ‘if/when she walks.’ In (47), triplicated *bwekúbweku* ‘gabbling, making a lot of noise’ is introduced by the locative relative clause *panak<sup>h</sup>araco* ‘where she is.’ In (48), *p<sup>h</sup>é:* acts as the predicate of *iyé k<sup>h</sup>oŋgo* ‘he the

impostor.’ Together, they form the main clause in which the temporal relative clause *pidafika iye* ‘when he arrives’ is embedded.

(46) *iyē kaŋkitikiti, aŋgafamba odzǐ-odzǐ-odzǐ-odzǐ*

[*iyē ka-ŋ-kiti-kiti*]<sub>INDPDT</sub> [*a-ŋga-famba*]<sub>SUB</sub>

PRO3SG 12(DIM)-1-small-RED SP1-SIT-walk

[*odzǐ-odzǐ-odzǐ-odzǐ*]<sub>MAIN</sub>

IDEO(walking/moving feebly, without strength)-RED

‘She is very small. When she walks, she walks feebly, without strength.’

(Conv.06\_#177)

(47) *ciŋkazi cire ciŋgabwera cisabwēka maŋiŋgi, panak<sup>h</sup>araco bwēkúbwēku-bwēkúbwēku-bwēkúbwēku*

[*ci-ŋ-kazi cire*] [*ci-ŋga-bwera*]<sub>SUB</sub> [*ci-sa-bwēka maŋiŋgi*]<sub>MAIN</sub>

7(DIM)-1-woman 7.DEM.III SP7-SIT-come SP7-HAB-gabble much

[*pa-na-k<sup>h</sup>ara = co*]<sub>SUB</sub> [*bwēkúbwēku-bwēkúbwēku-bwēkúbwēku*]<sub>MAIN</sub>

AGRP16-PRS-be.REL = PRO7 IDEO(gabbling)-RED

‘That woman, when she comes, she talks a lot of nonsense. Where she is, she makes a lot of noise.’ (Elicitation)

(48) *iyē k<sup>h</sup>oŋgo pidafika iyē p<sup>h</sup>é:: amona zǎ “nduyo porra azip<sup>h</sup>ura”*

[*iyē k<sup>h</sup>oŋgo*] [*pi-da-fika iyē*]<sub>SUB</sub> [*p<sup>h</sup>é::*]<sub>MAIN</sub>

PRO3SG 1a.impostor AGRP8-PFV.REL-arrive PRO3SG IDEO(observing, spying)

[*a-mu-ona zǎ*]<sub>INDPDT</sub> [*ndi-uyo porra a-zi-p<sup>h</sup>ura*]

SP1.PFV-OP1-see there COP-3SG damn! SP1-OP10-remove.from.fire

‘The impostor, when he arrived, observed/spied, and he saw him (the turtle):

“This is him, damn, he removed them from the fire!” (Story\_HareTurtle\_#49)

Following subordinate clauses, *odzǐ-odzǐ-odzǐ-odzǐ*, *bwēkébwēku* and *p<sup>h</sup>é::* express on their own an event which would otherwise be expressed by means of a main verbal clause. Their presence is syntactically required in the sentence.

In rare cases, ideophonic predicates do not seem to induce any argument structure, not even agent. In (49), for instance, the ideophone *cúŋwi* ‘splashing sound’ appears at the sentence margin following the description of the bean falling into a well. It



appears as a fully autonomous and argument-less predicate (also see example (51b) below for another illustration).

(49) *nyimo ire yap<sup>h</sup>unyuzika ŋk<sup>h</sup>ugwa muŋcera mure | cúbwi!*

[*nyimo ire ya-p<sup>h</sup>unyuzika*]<sub>MAIN</sub> [*ŋk<sup>h</sup>u-gwa mu-ŋ-cera mure*]<sub>SUB</sub>

9.bean 9.DEM.III SP9.PFV-slip COP15-fall 18-3-well 3.DEM.III

[*cú**ɓ**wi*]<sub>INDPDT</sub>

IDEO(splash)

‘That bean slipped, fell into that well, splash!’ (Story\_HareTurtle\_#108)

#### 6.4 Summary

Ideophones do not seem to display strong syntagmatic restrictions. They can all co-express and/or specify the event expressed by a verb, or form complex predicates with copulas and light verbs. They can also all be used holophrastically or predicatively: no restrictions seem to apply, at least as far as the possibility of integrating subject arguments is concerned. The possible presence of object arguments, in turn, is conditioned by the lexical properties of ideophones. Examples (50)-(51) illustrate multiple syntagmatic uses of ideophones in Sena. In (50), the ideophone *oyó-oyo* ‘speaking simultaneously’, repeated twice, can be used predicatively following the copula verb *ri*, or the light verb *cita*. There is no difference in meaning between the two constructions.

(50) *ndagumana ant<sup>h</sup>u ari/akucita oyó-oyo oyó-oyo*

*nda-gumana a-nt<sup>h</sup>u a-ri / a-ku-cita oyó-oyo-oyó-oyo*

SP1SG.PFV-meet 2-people SP2-be SP2-PROG-do IDEO(speaking

simultaneously)-RED

‘I met people talking all at the same time.’ (Elicitation)

In (51), the ideophone *bzwó* ‘dawning’ is used twice. In (51a), it is the complement of the light verb *cita* ‘do’ inflected for the first person singular, which marks the speaker as the agent of the event ‘waking up’. In (51b), *bzwó* is used holophrastically to refer to the time of daybreak.

(51) *ndingacita bzwó kunendene ηkwenderatu, bzwó kunendene ηkwenderatu*

a. *ndi-ηga-cita bzwó ku-na-enda = ine ndi-kwendera = tu*  
 SP1SG-SIT-do IDEO(dawning) AGRP17-PRS-go.REL = PRO1SG SP1SG-go.to = EMPH  
 ‘When I wake up, wherever I go, I do go.’ (LifeStory\_CeTh\_#69)

b. *bzwó ku-na-enda = ine ndi-kwendera = tu*  
 IDEO(dawning) AGRP17-PRS-go.REL = PRO1SG SP1SG-go.to = EMPH  
 ‘At dawn, wherever I go, I do go.’ (LifeStory\_CeTh\_#70)

## 7. Semantic aspects

This section is devoted to the semantic aspects of ideophones, addressing their general meanings (§7.1), the question of iconicity (§7.2), and the existence of idioms with ideophones and body parts (§7.3).

### 7.1 General meanings

Ideophones in Sena cover a large array of meanings that cover the five major semantic fields hierarchically identified by Dingemans (2012: 663) along an implicational hierarchy as follows: Sound > Movement > Visual pattern > Other sensory percepts > Inner feelings and cognitive states. This hierarchy makes predictions about the existence of members in each field: if a language has ideophones expressing movement, then it also has ideophones for sound, etc. These five semantic categories are very broad and could easily be split into various sub-fields. Sound evokes any sensation perceived by the sense of hearing, and includes onomatopoeic sounds such as animal voices, wind blowing, etc. Movement illustrates an act or process of moving, any physical behavior which implies a change of place or position or posture, or a particular instance or manner of moving. Visual patterns cover a range of physical properties such as color, form, dimension, aspect, state, etc. The other category of sensory perception involves tactile, gustatory, and olfactory sensory domains. Examples of each category are listed in (52). Attention has been given to provide examples uncited so far.

(52) a. Sound

<i>jigú-jigú</i>	‘sound of drawing water at the pump’
<i>gegebede</i>	‘noisy sound of something falling’
<i>gwidjó</i>	‘sound while eating or swallowing’

b. Movement

<i>jidú-jidú</i>	‘dancing in circles’
<i>c<sup>h</sup>ó-c<sup>h</sup>ó-c<sup>h</sup>ó</i>	‘jumping on one leg’
<i>nɣú-nɣú-nɣú</i>	‘hunkering down’

c. Visual pattern

<i>doríró:</i>	‘being clean and transparent (water)’
<i>dzá:</i>	‘being very tall, giant’
<i>bití-biti</i>	‘shining, glittering (skin, wood)’

d. Other sensory perceptions

<i>ǰé:</i>	‘spicy (food or drink)’
<i>gú:</i>	‘taste of sour food’
<i>rebvú-rebvú</i>	‘(too) sweet’
<i>k<sup>h</sup>ámwa</i>	‘wet’
<i>nyanvúrí-vuru</i>	‘getting goose bumps’

e. Inner feelings and cognitive states

<i>dodoméké</i>	‘being doubtful’
<i>dúmé:</i>	‘being afraid, with fear’
<i>nóni</i>	‘alone, lonely’
<i>nyandú-nyandu</i>	‘madly agitated’

Each of the semantic fields represented in (52) contains a large inventory of ideophones.

Deverbal ideophones usually display a one-to-one correspondence with the semantics of the verbs they are derived from. This explains why my main consultant was able to assign a specific meaning to many ideophones found in Prieto (2015) outside of their context. Postulating that the meaning of an ideophone depends on the context and is highly influenced by socio-pragmatic factors (Moshi 1993) is thus not entirely correct in Sena. Other ideophones do not have a clear connection to members of other word classes. Ideophones of onomatopoeic origin are proportionally very few in Sena, defying Doke’s (1935: 118) famous statement on the onomatopoeic properties of ideophones.<sup>8</sup>

A few ideophones are polysemous, as can be seen in (53).

---

<sup>8</sup> In Doke’s (1935: 118), an ideophone is “a vivid representation of an idea in sound. A word, often onomatopoeic, which describes a predicate, qualificative or adverb in respect to manner, color, smell, action, state or intensity.”

- (53) *dúr::* i) ‘silent with fear (in front of chiefs)’  
 ii) ‘fading away (for plant)’  
*p<sup>h</sup>urúr::* i) ‘unrolling mat’  
 ii) ‘trying to fly away, for trapped bird’  
*tímḍwí-tímḍwi* i) ‘with tightened belt’  
 ii) ‘suffering because of someone’  
*p<sup>h</sup>á-p<sup>h</sup>á-p<sup>h</sup>á* i) ‘move and sound of bird landing’  
 ii) ‘when crop starts growing and fields get green’

## 7.2 Iconicity in ideophones

In this subsection, I examine how certain phonological properties of ideophones (particularly deverbal ideophones) may correlate with the manner in which an action is performed. The shape of many deverbal ideophones follows two complementary patterns ruled by phono-semantic principles. The first pattern involves vowel lengthening, resulting in CV:(::) or CV:(::)CV shapes. Consonant lengthening is also attested, albeit rarer and restricted to the trill /r/ (see Table 6). The prolongation of a sound is an expressive means of describing a single, durative and usually intense event which may be extended in space or time. The ideophone *fó::* (< *kufota* ‘to fart silently’) expresses a single occurrence of the act of farting silently. The second pattern relies on reduplication, triplication and further repetition of an ideophonic basis, with the semantic effect of expressing a recurrent, repetitive or even rhythmic event. Onomatopoeic ideophones such as *wú-wu-wú* ‘sound of dog barking’ or *mé::-mé::* ‘sound of goat bleating’ are vivid examples. More examples are provided in (54), in which ideophones are paired according to these two phono-semantic principles. Notably, the triplication and even further repetition of monosyllabic ideophones, as *pé-pé-pé* in (54a) and *t<sup>h</sup>é-t<sup>h</sup>é-t<sup>h</sup>é* in (54b), generally applies to events made up of small actions (here short sounds) which repeat over and over.

- (54) a. *pé::* ‘sound of a long and continuous breath/blow’  
*pé-pé-pé* ‘sound of a repetitive gesture which involves air movement, such as waving a fan’  
 b. *t<sup>h</sup>é::* ‘sound of soft and continuous wind blowing’  
*t<sup>h</sup>é-t<sup>h</sup>é-t<sup>h</sup>é* ‘sound of soft wind blowing repetitively’  
 c. *nyá::ŋgu* ‘licking’  
*nyanḡú-nyanḡú* ‘licking repeatedly’

- d. *p<sup>h</sup>urú:* ‘trying to fly away (for trapped bird)’  
*p<sup>h</sup>urú-p<sup>h</sup>uru* ‘trying repeatedly to fly away (for trapped bird)’

In (55)-(56), ideophones derived from the verb stem *godoma* ‘kneel’ are illustrated. In (55), *gó::da* follows the copula verb and the event described, a woman on her knees, is static and durative. In (56), the reduplicated *godá-goda* supports the repetitive and rhythmic motion implied in the verb *kufamba* ‘to walk’.

- (55) *maye ure taŋgumana, ari gó::da*  
*maye ure ta-ŋ-gumana a-ri gó::da*  
 1.mother 1.DEM.III SP1PL.PFV-OP1-meet SP1-be IDEO(kneeling)  
 ‘That mother, we met her (she is) kneeling.’ (Elicitation)

- (56) *kufamba godá-goda*  
*ku-famba godá-goda*  
 INF-walk IDEO(kneeling)-RED  
 ‘To walk on the knees.’ (Elicitation)

These phono-semantic principles, according to which the shape of ideophones mirrors the physical characteristics of the actions they refer to, are in fact strongly linked to the so-called “diagrammatic iconicity” (Downing & Stiebels 2012), which establishes visual resemblance between the form of the sign or linguistic expression and its meaning.

In terms of sound symbolism, only a few (possible) correspondences between sound and meaning could be detected. The most promising example is the (aspirated) voiceless bilabial stop /p<sup>h</sup>/, which is associated with air movement in at least three onomatopoeias: *pé:* ‘sound of a long and continuous breath/blow’, *pé-pé-pé* ‘sound of a repetitive gesture which involves air movement, such as waving a fan’, and *p<sup>h</sup>á-p<sup>h</sup>á-p<sup>h</sup>á* ‘move and sound of bird landing’. The use of the aspirated voiceless velar stop /k<sup>h</sup>/ in *k<sup>h</sup>osó-k<sup>h</sup>oso* ‘sound of coughing a lot’ evokes abrupt vocalization. The aspirated voiceless alveolar stop /t<sup>h</sup>/ at the beginning of onomatopoeias correlates with a soft or quiet sound, e.g. *t<sup>h</sup>ó-t<sup>h</sup>ó-t<sup>h</sup>ó* ‘sound of drops of water dripping’, *t<sup>h</sup>é-t<sup>h</sup>é-t<sup>h</sup>é / t<sup>h</sup>é:* ‘sound of soft wind blowing repetitively/once’, and *t<sup>h</sup>úkutu* ‘landing gently’. These examples are too few, however, to speak of phonesthemes, and there is plenty of data that counters the above suggestions.

As for vowels, they do not seem to be affected by sound symbolism. No direct meaning may easily be associated to a vocalic sound. Furthermore, the process whereby vocalic variation triggers variation in meaning does not exist in Sena ideophones. This is illustrated in the minimal pairs in (57) based on vowel substitution. No systematic sound-meaning correspondences have been detected for vowels. The pairs of ideophones in (57) differ only in the quality of their vowels, and although a semantic connection might be detected, namely different sound types in (57a), alignment in (57b), and loss of order in (57c), the choice of vowels in each pair seems arbitrary.

- (57) a. *dré-dré* ‘sound of fart’  
           *dró-dró* ‘sound of jumps made by medium-sized animals (e.g. goat)’  
       b. *derú-derú* ‘balancing’  
           *durú-duru* ‘in a line or in a group (people walking)’  
       c. *dzedzé-dzedzé* ‘losing balance, staggering’  
           *dzudzú-dzudzú* ‘causing confusion’

### 7.3 Ideophone-body part collocations

Several ideophones in Sena combine with body part nouns. Two body parts are particularly conducive to attracting ideophones, i.e. *maso* ‘eyes’ (58) and *ɲtima* ‘heart’ (59).

#### (58) *maso* ‘eyes’ + ideophone

- maso goma-goma* ‘with a look of pain’  
*maso gudú-gudú* ‘with closed or heavy eyes (because of drowsiness)’  
*maso muni-muni* ‘with half-closed eyes’  
*maso t<sup>h</sup>ubu-t<sup>h</sup>ubu* ‘with swollen eyes’  
*maso cetú-cetu* ‘with threatening eyes’ or ‘with bright eyes’  
*maso mbú:* ‘with staring eyes’  
*maso rwá-rwá* ‘with eyes wide open, without blinking’

#### (59) *ɲtima* ‘heart’ + ideophone

- ɲtima bakú-bakú* ‘with heart beating’  
*ɲtima parú-paru* ‘with heart beating in fear’  
*ɲtima bii* ‘with fearful heart, with apprehension’

- ntima gugú:du* ‘with dry heart, with no empathy/compassion’  
*ntima répe-répe* ‘with the desire to do or eat sth.’  
*ntima dyókó-dyoko* ‘with the desire to eat sth.’

Often, the ideophone directly follows the body part noun (60), albeit a possessive modifier can easily be inserted between both elements (61).

- (60) *taṅgumana pakati mbakabvina maso rwá-rwá*  
*ta-ṅ-gumana pa-kati mba-ka-bvina ma-so*  
 SP1PL.PFV-OP1-meet 16-within CVB-PST.IPFV-dance 6-eye  
*rwá-rwá*  
 IDEO(eyes wide open)  
 ‘We found him here dancing, his eyes (are) wide open (in trance).’ (Elicitation)

- (61) (*wandipañgiza mano ako, mano ako niṅga ukaka,*)  
*maso ako cetú-cetu, kuyetima niṅga nyenyezi*  
*ma-so ako cetú-cetu ku-yetima niṅga nyenyezi*  
 6-eye AGRP6.POSS2SG IDEO(bright) INF-shine like 10.star  
 ‘(You showed me your teeth, your teeth like sesame seeds,) your eyes are bright, shining like stars.’ (Poem\_Ndiri.Ncitulo)

Syntactically, ideophones that take body parts subjects can be predicates on their own (see §6.3), as in (60) and (61) above, or form a complex predicate with a copula (see §6.2), as in (62) below, which literally translates as ‘my heart was with the desire to eat something’.

- (62) *papaya pikayionene ntimaṅga uk<sup>h</sup>ari répe-répe funa kuyidya*  
*papaya pi-ka-yi-ona = ine*  
 9.papaya AGRP8-PST.IPFV-OP9-see.REL = PRO1SG  
*ṅ-timaṅga u-k<sup>h</sup>a-ri répe-répe funa*  
 3-heart.POSS1SG SP3-PST.IPFV-be IDEO(desire to eat sth.) want  
*ku-yi-dya*  
 INF-OP9-eat  
 ‘The papaya, when I saw it, I felt the desire to eat it, I wanted to eat it.’(Elicitation)

Two other body parts are found in idioms with ideophones, i.e. *ṅk<sup>h</sup>ope* ‘face’ (63) and *mimba* ‘belly, stomach’ (64), but these combinations are less productive. Further note

that *yezi-yezi* ‘out of control like a drunk’ and *zígu* ‘with sudden gesture shifting away from someone’ in (63) express movements or gestures which are not restricted to the face; they also naturally combine with the noun *manungo* ‘body’.

(63) *ɲk<sup>h</sup>ope* ‘face’ + ideophone

<i>ɲk<sup>h</sup>ope gada-gada</i>	‘with big and/or open face’
<i>ɲk<sup>h</sup>ope nyasi-nyasi</i>	‘with upset face ready to cry’
<i>ɲk<sup>h</sup>ope yára-yára</i>	‘with false joyful facial expression while feeling resentful’
<i>ɲk<sup>h</sup>ope yézí-yézí</i>	‘out of control like a drunk’
<i>ɲk<sup>h</sup>ope zígu</i>	‘with sudden gesture shifting away from s.o.’

(64) *mimba* ‘belly, stomach’ + ideophone

<i>mimba mba-mba-mba</i>	‘with full stomach’
<i>mimba dé:fu</i>	‘with empty stomach’
<i>mimba toro</i>	‘swelling of pregnant belly’

Example (65), extracted from naturalistic data, illustrates the use of *mimba toro* ‘swelling of pregnant belly’.

(65) *pakusemba mai uyu, mwana ne nensa di mimba toro, mwana jí:*

<i>pa-ku-semba</i>	<i>mai</i>	<i>uyu</i>	<i>mwana</i>	<i>ne</i>	<i>nensa</i>	<i>di</i>	<i>mimba</i>
16-INF-marry	1.mother	1.DEM.I	1.child	NEG	be.difficult	COP	9.belly
<b><i>toro</i></b>			<i>mwana</i>	<i>jí:</i>			

IDEO(swelling of pregnant belly) 1.child IDEO(standing straight, well mounted)  
 ‘When this woman got married, the child, it was not difficult to get pregnant,  
 the child is now here standing.’ (Conv.06\_MaBeCe\_#343)

## 8. Conclusion: Toward a definition of ideophones in Sena

From the features discussed above, we first conclude that Sena ideophones may assume functions or distributions typically conveyed by other word classes. They have the ability to behave like adverbs or converbs, both in terms of distribution (mostly sentence-final) and function (descriptive or manner adverbs). This is especially the case when they combine with verbs from which they are morphologically derived, resulting in constructions that literally translate as ‘cut cutting’. Together with copula verbs, ideophones form complex predicates, e.g. *ri bídó:* ‘be dark, black’. In this environment, the syntactic behavior of ideophones shows more affinities with that of



nouns and adjectives, although ideophones do not share the morphological properties displayed in nouns and adjectives. In other syntactic environments, ideophones are used predicatively on their own, that is they substitute the verb and take thematic arguments. Unlike normal verb forms though, ideophonic predicates do not require concordial agreement with the nominals with which they pattern. Nor do they contain any type of inflectional or derivational morphology. Finally, they behave like interjections in the sense that they both express emotions, are easily lengthened or repeated with a characteristic prosodic pattern, and overall, they are phonologically and syntactically less constrained than other word classes. However, what clearly distinguishes ideophones from interjections is the capacity of the former to participate in the clause construction and have arguments.

What makes a lexeme an ideophone in Sena? Among the different word classes in the language, ideophones are distinctive in their phonology, their morphology, and above all their syntax. First, vowel (and more rarely consonant) length — in a system where long vowels are non-phonemic — is a defining property of Sena ideophones. Second, only the category of ideophones contains elements with closed syllables. Third, in a prosodic system marked by predictable stress, ideophones exhibit tone patterns. Morphologically, unlike most lexical word classes, including nouns, adjectives, verbs and adverbs, ideophones have no (or very little) morphology. The high frequency at which reduplication is applied in ideophones is also not seen in other word classes, excepts adverbs in *ci-* (§2). Triplication and multiplication, commonly found in ideophones, are not attested elsewhere, with the exception of interjections. Syntactically, ideophones can display a variety of functions, as explained in the preceding paragraph, but a crucial aspect of their syntactic behavior which distinguishes them from all other word classes - except interjections - is their capacity to be used independently and holophrastically.

In light of the above elements, Sena ideophones can be deemed as a distinct word class in the language on a par with nouns, adjectives, verbs, adverbs, etc. This is in line with the analyses usually adopted by Bantuist scholars, starting with Doke (1954) for Southern Bantu languages.

### **Acknowledgements**

This research has been carried out as part of my participation in the research group *Ideophones and interjections in a typological, areal and diachronic perspective* led by

Aimée Lahaussois and Yvonne Treis (LabEx *Empirical Foundations of Linguistics*, ANR-10-LABX-0083), who are gratefully acknowledged. I also wish to thank the two anonymous reviewers as well as Yvonne Treis and Aimée Lahaussois for their very helpful and constructive comments on an earlier version of this chapter. Parts of the research presented herein were supported by grants from the CNRS InSHS (“Soutien à la Mobilité Internationale” program) and the IFAS (French Institute of South Africa) for fieldwork data collection. I take this opportunity to warmly thank my Sena language consultants for their help and patience.

## Abbreviations

1, 2, 3... = noun class	FV = final vowel	PRO = pronoun
1SG/PL, 2SG/PL = grammatical persons	HAB = habitual	PROG = progressive
.I, .II, .III = demonstrative series (.I ‘proximal’, .II ‘medial’, .III ‘distal’)	IDEO = ideophone	PRS = present
AGRP = agreement prefix	INF = infinitive	PST = past
AM = associated motion	INT = intensive	REL = relative
APPL = applicative	INTERJ = interjection	RED = reduplication
COM = comitative	IPFV = imperfective	SEQ = sequential
CONN = connective	ITV = itive	SG = singular
COP = copula	NARR = narrative	SIT = situative
CVB = converb	NEG = negation	SP = subject prefix
DEM = demonstrative	NTR = neuter	VOC = vocative
DIM = diminutive	OP = object prefix	VTV = ventive
EMPH = emphatic	PASS = passive	
	PFV = perfective	
	PLA = plural addressee	
	POSS = possessive	

## References

- Akita, Kimi & Prashant Pardeshi (eds.). 2019. *Ideophones, mimetics and expressives*. Amsterdam: John Benjamins.
- Alves, Albano Padre. 1939. *Noções gramaticais da língua chisena*. Braga: Missões Fransiscanas.
- Ameka, Felix K. 2001. Ideophones and the nature of the adjective word class in Ewe. In F. K. Erhard Voeltz & Christa Kilian-Hatz (eds.), *Ideophones*, 25–48. Amsterdam: John Benjamins.

- Anderson, W.G. 1897. *An introductory grammar of the Sena language*. London: Society for promoting Christian knowledge.
- Andrason, Alexander. 2017. The “exotic” nature of ideophones - From Khoekhoe to Xhosa. Part 1. *Stellenbosch Papers in Linguistics* 48. 139–150.
- Andrason, Alexander. 2020. Ideophones as linguistic “rebels” - The extra-systematicity of ideophones in Xhosa. Part 1. *Asian and African Studies* 29(2). 119–165.
- Childs, G. Tucker. 1994. African ideophones. In Leanne Hinton, Johanna Nichols & John J. Ohala (eds.), *Sound symbolism*, 178–204. Cambridge: Cambridge University Press.
- Childs, Tucker G. 2003. *An introduction to African languages*. Amsterdam: John Benjamins.
- Cole, Desmond T. 1955. *An introduction to Tswana grammar*. London: Longmans Green.
- Courtenay, Karen R. 1976. Ideophones defined as a phonological class: The case of Yoruba. In Larry M. Hyman, Leon C. Jacobson & Russell B. Schuh (eds.), *Papers in African linguistics in honor of W. M. Welmers.*, 13–26. Los Angeles: University of California.
- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6(10). 654–672.
- Dingemanse, Mark, Damián E. Blasi, Gary Lupyan, Morten H. Christiansen & Padraic Monaghan. 2015. Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences* 19(10). 603–615.
- Doke, Clement M. 1931. *A comparative study in Shona phonetics*. Johannesburg: University of the Witwatersrand Press.
- Doke, Clement M. 1935. *Bantu linguistic terminology*. New York: Longmans, Green & Co.
- Doke, Clement M. 1954. *The Southern Bantu languages*. London: Oxford University Press.
- Downing, Laura J. 2019. Tumbuka prosody: Between tone and stress. In Emily Clem, Peter Jenks & Hannah Sande (eds.), *Theory and description in African linguistics: Selected papers from the 47th Annual Conference on African Linguistics*, 75–94. Berlin: Language Science Press.
- Downing, Laura J & Barbara Stiebels. 2012. Iconicity. In Jochen Trommer (ed.), *The morphology and phonology of exponence*, 379–426. Oxford: Oxford University Press.
- Fortune, George. 1961. *Ideophones in Shona*. London: Oxford University Press.
- Funnell, Barry J. 2004. *A contrastive analysis of two standardized varieties of Sena*. Pretoria: University of South Africa MA dissertation.

- Grégoire, Claire. 1975. *Les locatifs en bantou*. Tervuren: Royal Museum for Central Africa.
- Güldemann, Tom. 1999. The genesis of verbal negation in Bantu and its dependency on functional features of clause types. In Jean-Marie Hombert & Larry M. Hyman (eds.), *Bantu historical linguistics: Theoretical and empirical perspectives*, 545–587. Stanford: CSLI.
- Güldemann, Tom. 2008. *Quotative indexes in African languages. A synchronic and diachronic survey*. Berlin: De Gruyter Mouton.
- Güldemann, Tom & Ines Fiedler. 2019. Niger-Congo “noun classes” conflate gender with deriflection. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity: Volume I: General issues and specific studies*, 95–145. Berlin: Language Science Press.
- Guthrie, Malcolm. 1967. *Comparative Bantu: An introduction to the comparative linguistics and prehistory of the Bantu languages. Part I, Volume 1: The comparative linguistics of the Bantu languages*. Farnborough: Gregg Press.
- Haiman, John. 2018. *Ideophones and the evolution of language*. Cambridge: Cambridge University Press.
- Hinton, Leanne, Johanna Nichols & John J. Ohala (eds.). 1994. *Sound symbolism*. Cambridge: Cambridge University Press.
- Hulstaert, Gustaaf. 1938. *Praktische Grammatica van het Lonkundo (Lomongo)*. Antwerpen: De Sikkel.
- Instituto Nacional de Estatísticas. 2017. *IV recenseamento geral da população e habitação 2017*. Mozambique: Instituto Nacional de Estatísticas.
- Jespersen, Otto. 1965. *A modern English grammar on historical principles. Part VI Morphology*. London: George Allen and Unwin Ltd.
- Jones, Michael Allen. 1988. Cognate objects and the case filter. *Journal of Linguistics* 24. 89–111.
- Kilian-Hatz, Christa. 1999. *Ideophone: Eine typologische Untersuchung unter besonderer Berücksichtigung afrikanischer Sprachen*. Habilitation (postdoctoral) thesis, University of Cologne.
- Kiso, Andrea. 2012. *Tense and aspect in Chichewa, Citumbuka, and Cisena*. Doctoral dissertation, Stockholm University.
- Kunene, Daniel P. 1978. *The ideophone in Southern Sotho*. Berlin: Dietrich Reimer.
- Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: The University of Chicago Press.

- Meeussen, Achilles E. 1959. *Essai de grammaire Rundi*. Tervuren: Musée Royal du Congo Belge.
- Meeussen, Achilles E. 1967. Bantu grammatical reconstructions. *Africana Linguistica* 3. 79–121.
- Meinhof, Carl. 1912. *Die Sprache der Duala in Kamerun*. Berlin: Verlag von Dietrich Reimer.
- Moreira, Alexandre. 1924. *Practical grammatical notes of the Sena language*. St Gabriel-Mödling (Austria): Anthropos.
- Moshi, Lioba. 1993. Ideophones in KiVunjo-Chaga. *Journal of Linguistic Anthropology* 3(2). 185–216.
- Nurse, Derek & Gérard Philippson (eds.). 2003. *The Bantu languages*. London: Routledge.
- Prieto, Fernando Pérez. 2015. *Elementos da língua chisena e da cultura do povo de Sena*. Balaka (Malawi): Sociedade dos Missionários de África.
- Samarin, William J. 1971. Survey of Bantu ideophones. *African Language Studies* 12. 130–168.
- Schultze-Berndt, Eva. 2001. Ideophone-like characteristics of uninflected predicates in Jaminjung (Australia). In Erhard F. K. Voeltz & Christa Kilian-Hatz (eds.), *Ideophones*, 355–373. Amsterdam: John Benjamins.
- Steriade, Donca. 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* 5. 73–155.
- Torrend, Jules. 1900. *Grammatica do Chisena. A grammar of the language of the Lower Zambezi*. Chipanga Zambezia: Missão de Chipanga.
- Van de Velde, Mark, Koen Bostoen, Derek Nurse & Gérard Philippson (eds.). 2019. *The Bantu Languages*. 2nd edn. London: Routledge.
- Voeltz, F. K. Erhard & Christa Kilian-Hatz (eds.). 2001. *Ideophones*. Amsterdam: John Benjamins.
- Whiteley, Wilfred H. 1956. *A practical introduction to Gusii*. Nairobi: East African Literature Bureau.
- Williamson, Kay & Roger Blench. 2000. Niger-Congo. In Bernd Heine & Derek Nurse (eds.), *African languages: An introduction*. Cambridge: Cambridge University Press.
- Young, Thomas Cullen. 1932. *Notes on the speech of the Tumbuka-Kamanga peoples*. London: Religious Tract Society.

#### CONTACT

rozenn.guerois@cnrs.fr

# Ideophones and sound symbolism in Northern Amis (Austronesian)

ISABELLE BRIL

CNRS, UMR 7107 LANGUES ET CIVILISATIONS À TRADITION ORALE & LABEX EFL

Submitted: 10/01/2023    Revised version: 07/02/2024

Accepted: 08/02/2024    Published: 12/06/2024



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

## Abstract

This is a study of ideophones in Northern Amis, an East Formosan, Austronesian language of Taiwan. Ideophones depict sensory experiences, and they generally have the same phonological and phonotactic properties as other lexemes; however, some ideophones show consonant and vowel alternations denoting grades of intensity or pitch, and a number of ideophones denoting bad smells and tastes are characterized by specific sequences of phonemes. Crucially, beyond their purely depictive function, many ideophones are also used to describe sensory experiences, in which case, they are derived and inflected as verbs, modifiers and nouns; yet, their greater grammatical integration does not necessarily lead to the loss of their ideophonic properties. Rather, these two functions, depiction and description, are better seen as occurring on a continuum between two poles, as in Japanese (Dingemanse & Akita 2017). Only when the depictive function and the pairing of form and sensory depiction are lost do these lexemes become de-ideophonic, generally taking on non-sensorial semantics.

**Keywords:** Austronesian, ideophones, onomatopoeia, sound symbolism.

## 1. Introduction

Amis (ami) is one of the fourteen surviving Austronesian (also called Formosan) languages of Taiwan, spoken along the east coast. Northern Amis, one of the four main dialects, is spoken in the north-east around the city of Hualien. The other dialects

are Tavalong-Vataan, Central, and Southern Amis; they display significant differences in phonology, lexicon and morphosyntax (Tsuchida 1988), and according to native consultants, they are not immediately mutually intelligible.

Ideophones are an understudied topic in Formosan languages and in other Austronesian languages in general, leading Klamer (2001: 171) to write that Kambara (xbr; East Sumbanese, Indonesia) is “one of the few Austronesian languages with ideophones, or more accurately, with documented ideophones”. The most recent research on this topic in Western Austronesian languages is Lee’s (2017) on Seediq (trv; Formosan), and Rubino’s (2001) on Ilocano (ilo; Philippines). In his pioneering study of Austronesian roots, Blust (1988, 2009: 357, 2022) identified around 23% of CVC roots as being “onomatopoeic”; many of these roots have reflexes in Northern Amis (see Table 5 in the Appendix).

### 1.1. A brief typological outlook on Amis

Some features of Amis, restricted to information central for the understanding of ideophones, are briefly presented. Amis is a verb/predicate initial language with a complex voice system. Most lexical roots are categorially neutral, and most lexical categories (nouns, verbs, modifiers, etc.) are identified once lexical roots have been derived and inflected as stems in a given syntactic environment (Bril 2017). Common nouns and action nouns (like *a call*, *a jump*, etc.) are flagged by the noun marker *u*. Verbs are derived by voice-affixes that are selected in accordance with the prominent thematic role of the nominative argument, i.e., Actor Voice (AV) *mi-* if the nominative argument is an Actor; Non-Actor Voice (NAV) *ma-* for stative verbs whose nominative argument is an Experiencer or the locus of a property; Undergoer Voice (UV) *ma-* if it is a patient, etc. For instance, the root  $\sqrt{cudad}$  occurs as a noun *u cudad* ‘letter, book’ and as a verb derived for AV *mi-cudad* ‘study’;  $\sqrt{talem}$  occurs as a noun in *u talem* ‘a blade’, and as a stative verb derived as NAV in *ma-talem* ‘sharp’. Amis has a symmetrical voice system and bipartite alignment, the details of which are spelled out in Bril (2022). Constructions with AV *mi-* and NAV *ma-* select the absolutive alignment with nominative *k-* and a core argument case-marked by oblique *t-*; UV constructions have ergative alignment, with a nominative *k-* Undergoer and a genitive agent marked by *n-*.

Ideophones in Amis DEPICT perceptions and occur in their base form followed by *sa* ‘say’ or *han* ‘do so’. Some of them are also used to DESCRIBE perceptions, they are then derived and inflected as verbs, nouns, modifiers like other lexical roots, raising questions

as to whether they still count as ideophones; however, this grammatical integration does not necessarily lead to lost ideophonic properties, but is best seen as a continuum, the loss of ideophonic properties being concomitant with the complete loss of their basic depictive function.

### **1.2. A note on the glossing and translation of ideophones**

Some ideophones are glossed with a translation when they have a dedicated meaning, like *kapkap* ‘IDEO.grope (in the dark)’ in example (7) or *pikpik* ‘IDEO.flap (wings)’ in (9). Ideophones whose specific meaning is context-dependent are glossed IDEO along with their contextual meaning given in the translation line; one such case is *telaq* which depicts the sound of rifle shots, clapping, firecrackers (see example (29)). A translation is also given when an ideophone has a slightly different meaning as a verb and as a noun, like *kela~kelaq* in (28), which depicts a clattering noise and also refers to a tipcart producing such noise, it is glossed as IDEO.tipcart.

The paper proceeds as follows; the characteristic features of Amis ideophones and the differences with onomatopoeia are discussed in Section 2; some aspects of ideophones and their connection to sound symbolism are the focus of section 3, this includes ideophones with a templatic sequence of phonemes depicting bad smells and tastes; Section 4 teases apart the depictive vs. descriptive usages of ideophones possibly leading to lost ideophonic features. Northern Amis ideophones are put in a typological perspective in Section 5, and in a Formosan perspective in Section 6; Section 7 concludes.

## **2. The characteristic features of Amis ideophones**

Ideophones are an open lexical class and are “marked words that depict some sensory imagery” (Dingemanse 2011: 25). In Amis, these lexemes depict sounds and other perceptions. Amis speakers do not have a specific label for them; in the three existing dictionaries, which mostly bear on Central Amis (Fey 1986, Poinot & Pourrias 1966, Rata 2013), some ideophones are listed and translated as ‘having or making the sound of’. This study is based on a corpus of recorded oral productions,<sup>1</sup> completed by data collected with stimuli.<sup>2</sup>

---

<sup>1</sup> The corpus contains several hours of spontaneous oral productions (stories of various types and procedural texts), collected over approximately fourteen months of fieldwork in several villages in and around the city of Hualien.

<sup>2</sup> Some of McLean’s (2021) auditory stimuli were used, they are found at <https://osf.io/y9rk2/>.



## 2.1. Phonological, phonotactic and prosodic features of Amis ideophones

Ideophones use the same phonemic inventory of vowels (/i, ə, a, u/) and consonants<sup>3</sup> as other lexemes and contain no unusual phonemes. Like other lexemes, ideophones may have expressive vowel lengthening on their last syllable; this type of prosodic augmentation expresses duration or intensity as in *ca::s* [tsa::s]<sup>4</sup> ‘sizzling sound’, as well as emotional reactions or evaluation. Vowel lengthening also occurs on verbs, negative morphemes, quantifiers, adverbs, expressing degree and pragmatic emphasis or focus (Bril & Skopeteas 2021: 75-76).

Ideophones follow the same general phonotactic rules as other lexemes in Amis: there are no complex onsets, nor complex codas; yet, in fast speech, unstressed /ə/ may be dropped between a stop and an approximant consonant, resulting in restricted types of clusters, such as *t(e)ra~t(e)raq* [traɾáʔ<sup>h</sup>] ‘sound of gun fire’ (from *teraq* [təɾaʔ<sup>h</sup>]), *k(e)loh* ‘go thump’.<sup>5</sup> Vowel-initial and vowel-final words are not allowed, and consequently there are rule-inserted glottal stops in such positions, as in *?ahteng* ‘smell of feces, rotten meat’, *kurukuru?* ‘IDEO.tumble.down’. Word-internally, syllables can be open as in *keris* ‘scratch’ or closed as in *rasmus* ‘drizzle’. Ideophones follow the same stress rule as other words, with stress occurring on the last syllable of fully inflected words (Bril 2016, Bril & Skopeteas 2021).

Like all lexical roots, ideophones in their base form can be monosyllabic CVC or disyllabic CV(C)CVC, and they can be reduplicated. When the base and the reduplicated forms co-exist, reduplication is then signaled, as with *kung* depicting a single hollow knocking sound (of a big bell) and *kung~kung* for repeated sounds. Ideophones that do not occur as simple bases are written as one word, for instance *pikpik* ‘sound and movement of flapping wings’, or *kurukuru?* depicting the VISUAL effect of tumbling or rolling down. Reduplication is very common, and occurs at the left edge of all lexical roots with various grammatical functions: e.g., reduplicated CVC(V)~ stems used as verbs like *tela~telaq* ‘clap, crackle’ have aspectual, semelfactive or durative meanings, and CVC(V)~ reduplication of nouns has plural, collective or exhaustive meanings. In yet other cases, reduplication may have an expressive use.

<sup>3</sup> /ə/ is written <e>; /u/ has an allophone /o/ before velar /k/, epiglottal <q> /ʔ/ and pharyngeal fricative /h/; the glottal stop is written <?>.

<sup>4</sup> Lengthening is marked by colons [::].

<sup>5</sup> In Central Amis, *kloh* depicts the sound of something falling or a sound made in the throat (Rata 2013).

## 2.2. The contrastive features of ideophones and onomatopoeia

Ideophones and onomatopoeia are open lexical classes depicting sensory experiences. There are however differences between them.

### 2.2.1 Onomatopoeia

Onomatopoeia in Amis are strictly SOUND depicting; they can be used holophrastically like *bling*<sup>6</sup> in (1), depicting the sound of breaking glass, or like *pang* and *pes* in (2).

- (1) *ha! u tingalaway = aku! bling!*  
INTJ NM glass.pot = NOM.1SG ONOM  
“ah! my glass pot!” “bling!” (Frog story\_Ciyaw.020)<sup>7</sup>
- (2) “*pang pes!*”<sup>8</sup> *ma-paqtiw cira namaka puhung n-u ngabur!*  
ONOM ONOM UV-eject NOM.3SG from horn GEN-NM deer  
“thump splash!” He was thrown off the deer’s antlers!’ (into the water)  
(Frog story\_Ciyaw.057)

Most onomatopoeia are words of one or two syllables with the canonical syllabic CVC or CV(C)CVC structure; however, a few are less canonical and have a protracted vowel as in (3), or a protracted consonant, like *ssss* for a hissing sound. They also occur as direct speech acts followed by the quotative verb *sa* ‘say’ as in (3) and (4), where *tus* depicts the sound produced by a thrown object.<sup>9</sup>

- (3) *Wuuu sa k-ira wacu.*  
ONOM say NOM-DEIC dog  
“wooo” went the dog. (Frog story\_Ciyaw.019)

---

<sup>6</sup> Its non-canonical syllabic structure CCVC is due to the loss of an unstressed /ə/, *b(e)ling*.

<sup>7</sup> These are texts references under Toolbox.

<sup>8</sup> *Pes!* also depicts the sound of ritual wine-spraying with the mouth, performed by shamans to the gods. Ideophones and onomatopoeia in relation to shamanistic practices require further investigation.

<sup>9</sup> The use of *sa* as an “ideophone forming construction” is briefly mentioned in Central Amis by Wu (2006: 148).

- (4) *Tus sa = tu itini i ka~ka-butiq-an.*  
 ONOM say = PFV here LOC RDP~NFIN-sleep-NMZ  
 ‘It went “swish” here onto the bed.’ (Tatakulaq atu Hungti.029)

Another distinctive criterion is the fact that onomatopoeia are often repeated as prosodically distinct words of one or two syllables, like *tak tak tak* for the sound of rifles or machine-guns, or clattering sounds made by walking with clogs. Example (5) shows the repeated onomatopoeia *pungpung*.

- (5) *Pungpung pungpung pungpung sa = tu.*  
 ONOM say = PFV  
 ‘Pungpung pungpung pungpung” it went.’ (Tatakulaq atu Hungti (2).0058)

Other cases of repeated onomatopoeia are: *ki ki ki* for the sound of some birds; *kir kir kir* for the sound of small frogs and some birds; *yi yi yi* for the sound of dogs squealing; *ciya ciya ciya* for the sound of cicadas; *sing sing sing sing* for the sound of bells; *kiya kiya kiya* for the sound of rain; *tita tita* for the sound of clocks, of horse steps, of rain;<sup>10</sup> *tata tata* for the sound of people speaking; *kulu kulu kulu*, which depicts the sound of something falling into water as in (6), and a gurgling water or belly sound.

- (6) *Kulu kulu kulu sa-an a mu-lenek i tebun.*  
 ONOM say-LV LK AV-sink LOC well  
 ‘It went “glug, glug, glug” as it sank into the well.’ (Tatakulaq atu Hungti.029)

### 2.2.2 Ideophones

Out of the 4800 lexemes collected in my dictionary of Northern Amis, about 110 of them (4.4%) are identified as ideophones based on the following features. They are lexemes characterized by a strictly canonical CVC or CV(C)CVC syllabic structure, in contrast with some onomatopoeia that allow protracted vowels as in (3) or protracted consonants. Besides, if they are reduplicated, ideophones make up one prosodic word. Unlike onomatopoeia, ideophones are not holophrastic, and crucially, unlike onomatopoeia, their semantics is not restricted to sounds, they may depict motions or gestures producing a sound. When depictive, ideophones are followed by *sa* ‘say’

<sup>10</sup> Possibly a Chinese loan from 嘀嗒 (dídā) ‘tick tock’, 滴答 (dīdā) ‘sound of clock’.

or *han* ‘do so’ as in (7a), while in their descriptive use, they can be derived as verbs, for instance as an activity AV *mi-* verb in (7b), and more infrequently as nouns. Ideophones denoting VISUAL or TACTILE sensations tend to be followed by the light verb *han* ‘do so’ as in (7a).

(7)

- a. *Kapkap han n-ira.*  
IDEO.grope do GEN-3SG  
‘He groped (in the dark).’ (Arakakai.031, Ogawa)
- b. *Mi-kapkap cira.*  
AV-IDEO.grope NOM.3SG  
‘He groped (in the dark).’

In (8a) *hiwhiw* depicts a hissing or panting sound, while in (8b) it is derived into a stative *ma-* verb referring to a breathing condition and used descriptively.

(8)

- a. *Hiwhiw sa k-u suni n-u pa-hanhan n-ira.*  
IDEO say NOM-NM sound GEN-NM CAUS-breathe GEN-3SG  
‘IDEO went the sound of her breathing.’
- b. *Ma-hiwhiw cira.*  
NAV-IDEO NOM.3SG  
‘He has asthma.’

In (9a), *pikpik* ‘sound of flapping (wings)’ is used depictively, but in (9b) it is derived as an activity *mi-* verb describing a sound produced by a movement and used descriptively.

(9)

- a. *Pikpik han k-u sakubad.*  
IDEO.flap do NOM-NM wing  
‘Flap flap went the wings.’
- b. *Mi-pikpik t-u sakubad.*  
NAV-IDEO.flap OBL-NM wing  
‘(It) flaps its wings.’ (U Tawrayan a lutuk.057)

Not all ideophones can be used descriptively, and many are only depictive. Ideophones used descriptively are derived and inflected words that are more “grammatically integrated in the morphosyntactic structure of the utterance” (Dingemanse & Akita, 2017: 5-6). Akita (2017) and Dingemanse & Akita (2017) also point out the correlation between grammatical integration and the loss of ideophonic features. In Amis, however, this is a possible, but not necessary correlation. The criterion for assessing their retained ideophonic feature is the necessary pairing of form and sensory expression; only the loss of their depictive use with *sa* ‘say’ is evidence of de-ideophonization. This is developed in section 4.

### 2.2.3. Grey areas and continuum of usage between onomatopoeia and ideophones

To sum up, onomatopoeia are holophrastic, can be repeated as prosodically distinct words, while ideophones form a single prosodic word. However, there are a few such lexemes that are used as both onomatopoeia and ideophones. This is shown in (10a-b), the onomatopoeia *?aw ?aw ?aw* depicts the howl of dogs or human cries of pain, fear or distress, while in (10c), it is used as an ideophone, i.e., as one prosodic word with reduplicated CVC~CVC pattern, derived as an AV *mi-* verb with descriptive and expressive usage; they stand in contrast with non-ideophonic lexemes such as *ungar* ‘yell, shout’, *luwiq* ‘yell, squeal’, *libabuy* ‘bark’ which lack such sound-depicting effects.

(10)

- a. “*?aw ?aw ?aw*”    *sa = tu*    *k-u*    *wacu?*  
 ONOM                    say = PFV    NOM-NM    dog  
 ‘The dog went “aw aw”.’ (Mikungling tu *wacu*.027)
- b. “*?aw ?aw ?aw*”    *sa = tu*    *ci*    *Mayaw*.  
 ONOM                    say = PFV    PM    Mayaw  
 ‘“aw aw” Mayaw said’ (sound of complaint) (U nu tumuk a kungku.185)
- c. *Mi-?aw~?aw*    *k-u*    *wacu*.  
 AV-RDP~IDEO    NOM-NM    dog  
 ‘The dogs howled.’ (Mikungling tu *wacu*.027)

A similar flexible use occurs with *teraq* which is an onomatopoeia followed by *sa* in (11a), and is used as an ideophone in (11b) with partial CVCV~CVCVC reduplication; it occurs

as an ideophone with a descriptive function in (11c) where it is derived as an AV *mi*-verb, together with the verbalizer *sa*;- (iii) finally, *teraq* is used as a noun in (11d).

(11)

- a. *Teraq teraq teraq teraq tera::q sa k-u remes.*  
 ONOM ONOM:LENGT say NOM-NM blood  
 ‘drip drip drip drip dri::p went the blood.’ (Pangangan.059)
- b. *Tera~teraq sa k-u urad i sera.*  
 RDP~IDEO say NOM-NM rain LOC ground  
 ‘Splatter went the rain on the ground.’
- c. *Mi-sa-tera~teraq k-iyān u nanum.*  
 AV-do-RDP~IDEO NOM-DEIC NM water  
 ‘The water was splattering.’ (Age groups.052)
- d. *Si cacay a teraq n-u nanum.*  
 EXS one LK IDEO.drop GEN-NM water  
 ‘There was one drop of water.’

Despite this flexible range of uses, *teraq* preserves its sensory and expressive properties even when used descriptively. This is precisely what distinguishes it from non-ideophonic lexemes such as *l(e)saq* ‘leak, drip’ or *tesiq* ‘drop, drip’, which denote a process without sound depicting usage and which may not be repeated like *t(e)raq* in (11a).

Table 1 summarizes these distinctions. Onomatopoeia in Column 1 are used holophrastically or as direct speech acts followed by the quotative verb *sa*. Ideophones in column 2 depict a sensory experience, and are followed by *sa* ‘say’, or *han* ‘do so’. Column 3 contains ideophones that can be used both depictively and descriptively, the latter as more grammatically integrated words, i.e., as derived and inflected verbs or nouns like other lexical roots, and which retain their ideophonic properties. On the other hand, column 4 is the point where ideophones have faded or lost ideophonic properties, are even more grammatically integrated, for instance with modifying use. These also favor semantic extension towards non-sensory and more abstract meanings.

### 2.3 Ideophones with fossilized morphemes

A subset of ideophones contains fossilized formatives or infixes expressing intensity or more specific meaning.

1 ONOMATOPEIAS	2 IDEOPHONES	3 IDEOPHONES grammatically integrated	4 DE-IDEOPHONIC WORDS
are ONLY performed, or quoted by using <i>sa</i>	express sensory depiction, followed by light verbs <i>sa</i> or <i>han</i>	describe perception or a sensory event	are descriptive, with faded or lost ideophonic properties
holophrastic	NO derivation and MINOR grammatical integration	undergo derivation, grammatically integrated as nouns, verbs	highly grammatically integrated, derived and inflected as nouns, verbs, modifiers
prosodic foregrounding	prosodic foregrounding		
protracted sound, repetition, expressive lengthening	reduplication, (repetition), expressive lengthening	expressive lengthening	expressive lengthening
polysemy	restricted polysemy	restricted polysemy	semantic extension towards more abstract meanings

Table 1: Onomatopoeia and ideophones.

### 2.3.1 Ideophones containing fossilized formatives

Some ideophones contain a formative occurring as an initial syllable before the root. For instance, *siw* depicts a hissing sound produced by the wind or by some fast motion, and *bisiw*  $\approx$  *bisiw*<sup>11</sup> with the formative *be-*  $\approx$  *bi-* have a similar meaning and preserve the characteristic pairing of form and sensory depiction; they may appear in compounds like *ta bisiw* in (12).

- (12) *Ta bisiw han a tangasa = tu i lumaq.*  
 go IDEO do COMP arrive = PFV LOC house  
 ‘They went whiz straight back home.’ (Amis script.103)

<sup>11</sup> The symbol  $\approx$  stands for alternate forms.

Another case is the fossilized formative *te-*  $\approx$  *ti-* occurring on various ideophonic roots, for instance *kung* and *tekung*<sup>12</sup> ‘banging sound’, *bek* ‘thud’ and *tibek* ‘pound (rice)’, *paqpaq* ‘slap, clap’ and *tepaq* ‘slap, swat, strike’; *mirmir* ‘shiver, shake’ and *temir* ‘throb, beat, thump’ (heart) as in (13). Importantly, words with these formatives have a broader meaning, associating SOUND, GESTURE, VISION and, in the case of *temir*, INNER BODILY perception.

(13)

- a. *Mirmir sa k-u tireng.*  
 IDEO.shake say NOM-NM body  
 ‘Her body shivered.’ (Tatakulaq atu Hungti.0101)
- b. *Mi-sa-temi~temir k-u balucuq = aku.*  
 AV-do-RDP~IDEO.throb NOM-NM heart = GEN.1SG  
 ‘My heart is thumping.’

### 2.3.2 Infixed ideophones

Another subset of ideophones displays reflexes of PAN infixes \* <al>, \* <aR>, \* <aN>, whose reflexes in Amis are <al>  $\sim$  <ar> and <an>.<sup>13</sup> In Amis, <ar>  $\sim$  <al> express intensity, pluractionality and semelfactive actions, they mostly occur on reduplicated ideophonic CVC~CVC roots like *t<ar>uktuk* ‘beat, throb’, from *tuktuk* ‘sound of hitting, hammering, beating a drum’, which can also be used as a noun as in (14).

- (14) *U t<ar>uktuk n-u balucuq.*  
 NM <INFX> IDEO GEN-NM heart  
 ‘The pulse/throbbing of the heart.’ (the knockknock of her heart)

In Puyuma (pyu; Formosan), a few similar forms are mentioned by Teng (2008: 36-37) as denoting the property of an object. In Saisiyat (xsy; Formosan), Zeitoun (2023)

<sup>12</sup> Reconstructed as PAN \*tekuŋ<sub>2</sub> ‘bang; sound of a thud’ with a formative *te-* (Blust & Trussel, Austronesian comparative dictionary ongoing).

<sup>13</sup> Reid (1994: 330) hypothesizes the meaning of \* <aR> to be distributive or plural. Li & Tsuchida (2009: 359) suggest that \* <aN> might mean ‘having the sound or quality of’, as in Paiwan (pwn) and Puyuma (pyu; Formosan).



discusses fossilized infixes <al>, <ar> in onomatopoeic, sound denoting words. In Ilocano (Philippines), Rubino (2001: 310-311) mentions a similar <an> infix expressing intensity or the continuous aspect of an action, like *pirpir* ‘flutter’, *p<an>irpir* ‘continual fluttering’; *pekpek* ‘cram’, *p<an>ekpek* ‘resound’ (of wood when hit). According to Rubino (2001: 314-315), the formerly productive infix <ar> occurs in “lexicalized items associated with the sound of the word” like *togtog* ‘knock’ and *t<ar>ogtog* ‘repeated knocking’; *ng<ar>asngas* ‘crunch’; *ng<ar>ebngab* ‘gnash one’s teeth’; *ng<ar>etngat* ‘gnaw’, *b<ar>asabas* ‘sound of heavy rain’, *k<ar>apkap* ‘grope in the dark’.

Northern Amis has around fifteen such infixes ideophonic forms, some of which are fossilized and no longer occur in their root form. Among them are *h<an>inghing* ‘smell of toilets, of passing wind’; *k<ar>atkat* ‘coarse, rough’; *k<al>aqkaq* ‘rattling noise, speak like a machine gun’ as in (15); *s<al>ucsuc* or *s<ar>ucsuc* ‘fluent, fluid’ which applies to liquids or figuratively to speech; *ng<al>iwngiw* ‘mumble, grumble murmur’ as in (16); *b<al>ambang* ‘churning, chaotic movement’ and its sound.

- (15) *Kalaqkaq k-ira taw.*  
 IDEO NOM-DEIC person  
 ‘This person rattles like a machine gun.’

They are used with *sa* ‘say’ (16a), or derived as voice-affixed verbs as in (16b). Some of them may undergo metathesis, as in (16c). The non-ideophonic lexeme is *cihi* ‘scold’.

- (16)
- a. *Ngaliwngiw sa k-u wina.*  
 IDEO say NOM-NM mother  
 ‘The mother mumbled to herself.’ (Arikakay Dawa.025)
- b. *Ngaliwngiw-an n-i ina k-iya wawa.*  
 IDEO-LV GEN-PN mother NOM-DEIC child  
 ‘The child was scolded by mother.’
- c. *Ngaliwngiw* or *langiwngiw* ‘mumble, murmur’.

Some infixes ideophones show semantic specialization, like *t<an>ektek* ‘strong, steady, earnest’, from *tektek* ‘sound of hitting on wooden post’.

In some cases, the base form and the infixed form co-exist, for instance *beriw* that depicts the wind's sound and motion, is the infixed form of *biw* 'sound of wind' (infixated as *b<ar>iw*, with the unstressed vowel weakened to /ə/). *Biw* is only depictive, while *beriw* has both functions, (i) depictive, followed by *sa* 'say' in (17a), possibly reduplicated as in (17b), or (ii) descriptive, when derived as *mi-sa-beriw* 'to blow a breeze' itself derivable as the participial modifier *mi-sa-beriw-ay* of *bali* 'wind' in (17c).

(17)

- a. *Beri:w*<sup>14</sup>      *sa = tu*      *k-u*      *baliyus*.  
 IDEO:LENGT    say = PFV    NOM-NM    typhoon  
 'The typhoon went whoosh.' (U surit nu Pangcah.189)
- b. *B(e)ri~b(e)riw*<sup>15</sup>    *sa*    *k-u*      *bali*.  
 RDP~IDEO            say    NOM-NM    wind  
 'The wind went whizz whizz.'
- c. *Saqqiq k-ina*      [*mi-sa-b(e)ri~b(e)riw-ay a*] *bali*.  
 cool    NOM-DEIC    AV-do-RDP~IDEO-MODF    LK    wind  
 'This whiff of wind/breeze is cool.'

In all such cases, in their descriptive use, ideophones preserve their identifying pairing of form and sensory expression. Other cases of infixed ideophones are:

(18)

*c<al>ascas* or *ca<ar>ascas* 'loud noise of water fall', 'loud noise of voices' from *cas~cas* 'torrent, cascade' and *cas* 'sizzling sound' (which is conceptualized as a related sound);

*k<al>ungkung* 'loud sound' (of thunder, of demolishing, etc.) from *kung~kung* 'sound of knocking on (a door, drum)';

*s<al>angsang* 'loud noise of rain or water-fall', from *sang* 'pitter-patter, sound of rain';

*t<ar>aktak* 'sound of something falling one by one' from *taktak* 'sound of spilled, scattered object'.

<sup>14</sup> Vowel lengthening expresses emphasis.

<sup>15</sup> Pronounced [b(ə)rib(ə)'riw].

### 3. Ideophones with templatic features and sound symbolism

Since ideophones are symbolic representations of resemblances between linguistic form and sensory experiences, their iconicity is conventional, as is sound symbolism. Blust (1988: 57-58) pointed out “a high incidence of initial /k-/ or /g-/ in morphemes that refer to rubbing, scratching, scraping [...], and many initial /ŋ/ in words relating to the oral or nasal area”. Northern Amis has a fair number of lexemes with initial voiceless plosives /k, t, p/ for hitting, scratching, scraping sounds, sounds of thunder or guns; it has lexemes with initial /h/ which denote breathing, snoring, while those with initial /s/ denote hissing sounds. However, since these phonemes occur in non-ideophonic lexemes and without any clearly skewed ratio, investigation of this type of sound symbolism is not pursued. On the other hand, there are ideophonic<sup>16</sup> sets identified by Blust (1988: 27, 38) as having a similar meaning and a specific syllabic template with consonant and vowel alternations denoting grades of intensity or pitch. Such Gestalt symbolism is inherent to ideophones since they depict a perceptual event, with reduplication used to express iterativity and “the vowel space to express grades of intensity between related forms” (Dingemanse 2012: 659, 663). Northern Amis has such sets of ideophones.

#### 3.1. Ideophonic sets, sound symbolism and their semantic effects

A few ideophonic sets vary by their initial consonants. One such case is  $C_1iw$ , in which the initial consonant can be labial *biw*, sibilant *siw*, fricative *hiw*. All of them depict a sound generally produced by some fast whirling or spinning movement connected to air or wind, like *biw~biw* depicting the sound of wind gusts, *hiwhiw* depicting a hissing breathing sound, *riwriw* depicting a spinning movement and its sound (like a spintop).

Others, approximately 35 of them, display vowel alternations and/or voiced vs. voiceless consonant alternation. They mostly express grades of intensity and related shades of meanings, like *kuris* ‘scratch, itch’ and *k(e)ris* ‘scratch, rub, rustle’, which refer to an action and connected sound.

- (19) *Sa-k(e)ris~k(e)ris*    *k-u*        *urad*.  
           do-RDP~IDEO            NOM-NM    rain  
           ‘The rain makes a drizzling sound.’

<sup>16</sup> Blust (1988) calls them “onomatopoeic”.

### 3.1.1 The semantic effects of vocalic alternations in ideophones

In his discussion of patterns of vowel alternations in Proto-Austronesian roots, Blust (2009: 357) states that “/a/ denotes loud, raucous sound, /ə/ muffled or blunted sound, /i/ a high-pitched sound and /u/ a loud or deep sound”.

In Amis by contrast, vowel alternations with high back vowel /u/ denote louder and deeper sounds, while those with high front /i/ and central /ə/ vowels denote higher pitch or sounds made by lighter objects. Among them are:

(20)

*k(e)lung* ‘sound of slammed door’ vs. *k(e)le~k(e)leng* ‘clap of thunder’ vs. *k(e)ling*<sup>17</sup> ‘sound of bell, clock, phone’;

*kungkung* ‘sound of knocking’ (on door, drum) vs. *kingking* ‘ring, tinkle (like bells)’;

*pukpuk* ‘sound of husking (rice)’ (in a container) vs. *pikpik* ‘flapping sound’ (of wings);

*tuktuk* ‘sound of hitting’ (door, wood, drum) vs. *tektek* ‘sound of wooden post being hit’ (into the ground) vs. *tiktik* ‘sound of patting, tapping’.

Some associate sound and gestures like *tuhtuh* ‘empty & dump’ vs. *tihlih* ‘shake off, brush off’.

Vowels /a/, /i/, /ə/ also alternate as in (21), where some of these ideophones associate GESTURES and VISUAL experience WITH or WITHOUT SOUND:

(21)

*bahbah* ‘expel, throw away’ vs. *bihbih* ‘shake off, brush off’;

*cakcak* ‘hoe, dig and loosen the soil’ vs. *cekcek* ‘squeeze, peck, press’ vs. *cikcik* ‘chop, slice’;

*paqpaq* ‘slap, clap, pat’ vs. *piqpiq* ‘press, tread’;

*taptap* ‘hit with an adze, pickaxe, chisel’ vs. *teptep* ‘chopping sound, sound of boat engine’;

*tera~teraq* ‘pitter-patter, sound of rain, of water’ vs. *teri~teriq* ‘tapping sound’ (lighter, as on a door).

---

<sup>17</sup> The first syllable is a fossilized formative, the second one is reconstructed as PAN \*linj ‘clear ringing sound’ (Blust 1988: 121).

- (22) *Araw*      *teri~teriq*    *sa*    *a*      *tengil-an*    *n-ira*.  
 suddenly RDP~IDEO    say    COMP    hear-LV      GEN-3SG  
 ‘Suddenly, there was a knock knock that she heard.’ (Chen 1969: 29)

### 3.1.2 The semantic effects of consonant gradation in ideophones

Consonant gradation is a prominent feature of ideophones in Austronesian languages. Blust (1988: 44) remarked that initial voiced stops “generally signal a louder or deeper sound in onomatopoeic roots than the homorganic voiceless stops” and that “voiced consonants signal larger sound-producing objects than their voiceless equivalents” (Blust 2009: 357). In contrast, voiceless stops in Northern Amis denote greater intensity.

- #C- voiced vs. voiceless consonant gradation associating gesture and sound:

*pikpik* ‘flap (wings), flutter’ vs. *bikbik* ‘sway arms, shake’;

*pihpih* ‘fan’ vs. *bihbih* ‘shake off, brush off’;

*pekpek* [pəkpək] ‘pound, hammer’ vs. *bek* [bək] ‘thud’.

- Initial /k, t/ alternations:

*kelaq* ‘creak, rattle, clatter’ vs. *telaq* ‘clap, crackle’.

Some ideophones denoting sounds of hitting, scraping and scratching also have final consonant /t/, /s/, /h/ alternations, in addition to vowel alternations, like *kutkut*, *kitkit* ‘scrape, scratch’, *kiskis* ‘scrape, scratch’, *kihkih* ‘scrape’. Final consonant alternation expressing intensity involves the velar /k/, the epiglottal <q> /ʔ/, the pharyngeal fricative /h/, such as *tiktik* ‘pat (to extract)’, *tigtig* [tiʔtiʔ<sup>h</sup>] ‘strike against something (hard)’, *tihthih* ‘shake off (dust), brush off’. All such ideophones are followed by *sa* ‘say’ and express related meanings with intensity gradation.

### 3.2. Ideophones with specific templates and skewed frequency

Bad smell and taste words also evidence phonesthetic properties with paired form-and-meaning occurring with a skewed frequency (Bergen 2004: 293). These words have specific templates with sequences of phonemes occurring as their first syllable, **Caŋ**CVC for smells and **Ca<sub>h</sub><sub>2</sub>/l<sub>2</sub>(e)**CVC for tastes.

Amis speakers spontaneously identify and list them as forming a distinct set of expressive words for bad smells and tastes. This is in line with Bergen’s remark (2004: 303-304)

that language users internalize recurrent form-meaning correlations, “such generalizations over individual tokens are based not on compositionality or productivity, but rather on the simple recurrence in the language of a form-meaning association”.

Their restricted number, the fact that words with such semantics are in great majority (up to 80%) words with such template, and the fact that no other words share this pattern, with one or two exceptions, identify them as ideophones of a distinct type. The reason for this template could be the presence of a now unanalyzable semantic or ideophonic formative. However, there are differences between them, bad smell ideophones having depictive and descriptive functions, while taste words tend to lose their depictive function.

### 3.2.1 Smell ideophones

Many of the ideophones expressing stench and bad smells are reconstructed in PAN with the pattern \*(C)an(e)CV(C) (Blust 1988: 58-60). They have reflexes in several Formosan and in other western Austronesian languages.<sup>18</sup>

There are fifteen such words in Northern Amis, nine of them with initial glottal stop like *?angcep* ‘smell of burnt (rice, food)’, three of them with initial /b/. These words make up the great majority (75%) of smell words in my lexical database (see Table 2 below, and Table 6 in the Appendix). Only five other non-ideophonic words referring to smells do not follow this pattern, one of them is *?ahteng* ‘smell of faeces, of rotten meat’, another is the generic word *sanek* ‘odour, smell’.

<b>?anCVC template</b>	<b>banCVC template</b>
<i>?angcep</i> ‘smell of burnt (rice, food)’	
<i>?angcuh</i> ‘smell of urine’	<i>bangruh</i> ‘smell of rotten food’
<i>?anglih</i> ‘smell of cut grass’	<i>banglih</i> ‘bad smell of grass, tree’
<i>?anglis</i> ‘smell of fish’	
<i>?angliw</i> ‘smell or taste of rotting food’	
<i>?angsit</i> ‘smell of animal or dirty person’	
<i>?angsaw</i> ‘smell of fire smoke’	<i>bangsit</i> ‘stench of urine’
<i>?angtir</i> ‘smell of sweat or of women's periods’	
<i>?angtul</i> ‘stinky’	

**Table 2:** Bad smell ideophones in Amis.

<sup>18</sup> In Puyuma (pyu), Paiwan (pwn) both Formosan, and in Philippine languages (Blust 1988: 60).

When depictive, smell ideophones are followed by *sa* ‘say’ (23a); when used descriptively, they behave like stative predicates, and can be derived as modifiers (23b). They are gradable, and may be derived as attenuative predicates with the reduplicated CVC~root-*an* pattern (23c-d). When negated, they contradict or express disagreement with a previous statement.

(23)

- a. *?angtul sa a sanek-an.*  
 stinky say COMP smell-LV  
 ‘It is stinky to smell!’ (Kakunas.021)
- b. *?angtul-ay a ka-ka?en-en k-inian.*  
 stinky-MODF LK NFIN-eat-UV NOM-DEIC  
 ‘This is stinky food.’ (U misamaraday.010)
- c. *?ang~?angtul-an.*<sup>19</sup>  
 CVC.RDP~stinky-AN  
 ‘It was a little smelly.’
- d. *?ang~?angcep-an.*  
 CVC.RDP~smell.burnt-AN  
 ‘It slightly smelt of (something) burnt.’

### 3.2.2 Taste words

Tastes words are an even smaller set of words (fewer than ten), with a **Cah<sub>2</sub>/l<sub>2</sub>** (e)CVC template and /h/ or /l/ as C<sub>2</sub>, like *?ahbed* ‘chewy’, *?ahcid* ‘salty’, *?alsuq* ‘tasty’, *?al(e)ngel* ‘bitter’, *?al(e)dah* ‘spicy’. They make up 81% of all attested taste words, with only two other words having a different pattern, in addition to the generic verb *tanam* ‘taste’. They are used to depict a personal experience, followed by *sa* as in (24a). They can be derived as inchoative verbs expressing a change of state as in (24b), and they can be negated to express disagreement, as in (24c).

(24)

- a. *?ahbed sa k-ina titi.*  
 chewy say NOM-DEIC meat  
 ‘This meat is chewy.’

<sup>19</sup>The suffix -AN in the attenuative CVC~root-*an* construction is homophonous with LV -*an*.

- b. *Mala-ʔahbed = tu k-ina titi, tada ma-tenes k-isu*  
 INCH-chewy = COS NOM-DEIC meat INTENS NAV-be.long NOM-2SG  
*a mi-cacak.*  
 COMP AV-cook  
 ‘The meat has become chewy, you cooked it too long.’
- c. *Caay ka-ʔahbed.*  
 NEG NFIN-chewy  
 ‘(They) are not chewy.’ (Katacumuli.013)

Like bad smell words, they can be derived as modifiers as in (25a); they are gradable and derived as attenuative predicates (25b); they may also denote a caused change of state (25c-d).

(25)

- a. *ʔaldah-ay a nanum.*  
 spicy-MODF LK water  
 ‘alcohol’ (lit. spicy liquid/water)
- b. *ʔal(e)~al(e)dah-an k-ina sinabel.*  
 CVC.RDP~spicy-AN NOM-DEIC dish  
 ‘This dish is a little spicy.’
- c. *Pa-ka-ʔaldah cira t-u sinabel.*  
 CAUS-NFIN-spicy NOM.3SG OBL-NM dish  
 ‘(s)he made the dish spicy.’
- d. *Pa-ʔahcid k-u cilaq t-u kabi.*  
 CAUS-salty NOM-DEIC salt OBL-NM soup  
 ‘The salt has made the soup (too) salty.’ (Chen 1987: 260)

The taste word *ʔalsuq* ‘tasty’ (26a) also extends to music or songs as pleasant auditory experiences as in (26b).

(26)

- a. *ʔalsuq-ay a ka-kaʔen-en.*  
 tasty-MODF LK NFIN-eat-UV  
 ‘(they are) delicacies (lit. it’s tasty to be eaten).’



- b. *ʔalsuq a tengil-an k-u radiw n-ira.*  
 pleasant COMP listen-LV NOM-NM song GEN-3SG  
 ‘Her song is beautiful to listen to.’

To summarize, the words for bad smells and tastes make up a very small subset of the lexicon with their own specific templates and highly skewed distribution that identify them as a distinct set of expressive words. However, taste words are mostly used descriptively and are more frequently de-ideophonized than smell words.

#### 4. The depictive vs. descriptive use of ideophones

Amis is not unique in having descriptively used ideophones. Dingemanse (2017: 376-380), Dingemanse & Akita (2017: 6-7) and Akita (2017: 316) point out similar facts in various languages in Africa, Eurasia, Australia, North and South America. They point out an inverse correlation between expressiveness and grammatical integration, also observed cross-linguistically. Akita (2017: 317, 323) proposes the following scale of morphosyntactic integration of Japanese (jpn) ideophones:

acategorical < quotative-adverbial < bare-adverbial < verbal < nominal
NON-INTEGRATED <span style="float: right;">INTEGRATED</span>

In Amis, ideophones also evidence this inverse correlation. Even when derived, they retain a close semantic relation with sensory features, as shown by the fact that the same ideophone can often be used depictively or descriptively (columns 2 and 3 of Table 1).

Dingemanse (2023: 9, preprint) mentions that in Ewe (ewe; Ameka 2001) and Basque (eus; Ibarretxe-Antuñano 2017), ideophones are multicategorical with predicative and attributive uses, while remaining mostly free of inflectional or derivational morphology.

##### 4.1. The descriptive usage and grammatical integration of ideophones

In Amis, ideophones occur under their base or reduplicated form, followed by *sa* ‘say’ as in (27a) or by *han* ‘do so’ when used depictively, some of them can also be derived and inflected as verbs, nouns or modifiers, in which case they describe a sensory event in an expressive way. For instance, the ideophone in (27b) is derived as a sound

denoting verb in locative voice, a voice that is typically used for natural phenomena and weather events. Compare with the non-ideophonic lexeme *betiliq* ‘thunder’, also derived as a verb *betiliq-an* in (27c).

(27)

- a. *Kele~keleng sa k-u betiliq.*  
 RDP~IDEO say NOM-NM thunder  
 ‘The thunder is rumbling.’
- b. *Kele~keleng-an.*  
 RDP~IDEO-LV  
 ‘It’s rumbling.’
- c. *Betiliq-an anini.*  
 thunder-LV now  
 ‘It’s thundering now.’

Ideophones derived as voice-affixed verbs describe a sensory event in an expressive way, possibly conveying a personal reaction to that event, a property lacking in non-ideophonic lexemes with similar meanings. Ideophones have a variable potential for derivation and inflection as verbs, modifiers and nouns. However, such grammatical integration does not necessarily correlate with lost ideophonic status, as long as they retain some expressive sensory features of their original depictive function. Though ideophones in descriptive use mostly occur in affirmative constructions, some of them may be negated. This is not specific to Amis, as various other languages allow ideophones to be negated, like Japanese or Hausa (hau; Dingemanse 2017: 364; 2023, preprint pp. 9-10).

For instance, the ideophone *kelaq* is depictive in (28a); it is derived and inflected as a sound denoting verb in the prohibitive mood in (28b), and in (28c) it is used as a noun referring to a ‘wooden tipcart’ producing that sound.

(28)

- a. *Kelaq! sa.*  
 IDEO say  
 ‘It went “crack”!’ (Cabay aku.00085)
- b. *Aka pi-sa-kela~kelaq!*  
 PROH NFIN-do-RDP~IDEO  
 ‘Don’t make clattering noise!’ (don’t clomp clomp with your clogs)  
 (Chen stories 1969: 26)

- c. *U muli n-u kela~kelaq.*  
 NM wheel GEN-NM RDP~ IDEO.tipcart  
 ‘The wheels of the rattling tipcart.’

The ideophone *t(e)la~t(e)laq* depicts various related sounds, rifle shots, firecrackers as in (29a), clapping. In (29b), it is a derived verb in Actor Voice *mi-sa-t(e)lat(e)laq*; in (29c), it is derived as a verb in Locative Voice *tela~telaq-an*, and in (29d), it is a deverbial noun *ni-pi-tela~telaq* ‘clapping’. On the other hand, non-ideophonic lexemes *bakuhac* ‘explode’, *kemiq* ‘shoot’ or *telik* ‘explode like firecracker, glitter’ have no depictive use.

(29)

- a. *Tela~telaq sa k-u pangpawa.*  
 RDP~IDEO say NOM-NM ONOM.firecracker  
 ‘The firecrackers went crackle crackle.’
- b. *Mi-sa-tela~telaq k-u pangpawa.*  
 AV-do-RDP~IDEO NOM-NM ONOM.firecracker  
 ‘The firecrackers are crackling.’
- c. *Kapah k-u radiw n-ira,*  
 good NOM-NM song GEN-3SG  
 ‘Her song was nice,  
*manay tela~telaq-an n-u alumanay.*  
 so RDP~IDEO-LV GEN-NM people  
 so people clapped (for her).’
- d. *Ma-tiya u betiliq k-u ni-pi-tela~telaq<sup>20</sup>*  
 NAV-be.like NM thunder NOM-NM PFV.NMZ-NFIN-RDP~IDEO  
*n-u alumanay.*  
 GEN-NM people  
 ‘The clapping of the people sounded like thunder.’ (Age groups.197)

The grammatical integration of ideophones does not necessarily incur loss of their sensory features as shown by *kelaq* in (28), *t(e)la~t(e)laq* in (29) and by *tektek* in (30).

<sup>20</sup>AV *mi-t(e)la~t(e)laq* is nominalized by the perfective NMZ *ni-*, together with the non-finite *pi-* form.

In (30a), *tektek* depicts the sound of hitting on wooden posts, while in (30b), it is derived as a causative verb inflected for hortative mood and used descriptively.

(30)

- a. *Tektek sa k-u suni n-iya kilang a mi-rangat.*  
 IDEO say NM-NM sound GEN-DEIC wood COMP AV-fence  
 ‘Tektek goes the sound of the wood as (he) makes a fence.’
- b. *Pa-tektek-a k-ita a mi-rangat.*  
 CAUS-IDEO-HORT NOM-1PL.INCL COMP AV-fence  
 ‘Let's hit the sticks (into the ground) to make a fence.’

*Bengbeng* depicts a banging sound in (31a), while in (31b), it describes a sound caused by pelting stones. The corresponding non-ideophonic verb is *mi-alud* ‘throw stones, lapidate’.

(31)

- a. *Bengbeng sa k-u suni n-ira panan.*  
 IDEO say NOM-NM sound GEN-DEIC door  
 ‘Bangbang went the sound of the door.’
- b. *Pa-bengbeng k-uhni t-u bek(e)loh.*  
 CAUS-IDEO NOM-3PL OBL-NM stone  
 ‘They banged (it) with stones.’

The ideophone *bek* depicts a dull sound or thud in (32a), while in (32b), it is marked by the conveyance voice affix *si-* indicating a change of location of the patient pivot and describes a sound caused by hurling a frog (the patient being unexpressed).

(32)

- a. *Bek sa k-u bek(e)loh a ma-terak.*  
 IDEO say NOM-NM stone COMP NAV-fall  
 ‘The stone went thud when it fell.’
- b. *Si-pa-bek han=tu itira i tebun.*  
 CV-CAUS-IDEO do.SO = PFV there LOC well  
 ‘(It) was made to go thud there onto the well.’ (Tatakulaq atu Hungti (2).0126)

*Tekiq* followed by *sa* in (33a) depicts striking, hitting sounds, while in (33b), it is derived as a noun marked as the nominative argument *k-u teki~tekiq* of the negative existential verb *awaay* and used descriptively. In (33c), it is a deverbal noun derived from the voice-affixed verb *mi-sa-teki~tekiq* ‘make a clanking noise’, which refers to noises made by human beings, and metaphorically to life.

(33)

- a. *Tekiq sa i walu<sup>w</sup>-ay k-u tuki.*  
 IDEO say LOC eight-<sup>ep</sup>-MODF NOM-NM clock  
 ‘Cling went the clock at 8.’
- b. *Awaay k-u teki~tekiq n-ina niyaruq.*  
 NEG.EXS NOM-NM RDP~IDEO GEN-DEIC village  
 ‘There was no clank clank sound in the village.’ (i.e., no sound of life)
- c. *Awaay = tu a maka-tengil k-aku t-u*  
 NEG.EXS = PFV COMP ABILT-hear NOM-1SG OBL-NM  
*pi-sa-teki~tekiq n-ira.*  
 NFIN-do-RDP~IDEO GEN-3SG  
 ‘I would no longer hear any clanking (sound) of his.’ (i.e., he is deceased)  
 (Cabay aku.00167)

On the other hand, non-ideophonic verbs like *palu* ‘strike with a stick’ and *mukun* ‘to hammer’ only describe actions. Another difference is that instrument nouns can be derived from action verbs by the dedicated prefix *sa-*, like *sa-mukun* ‘a hammer’, while sound-depicting ideophones are not derived as sound producing instruments by *sa-*. Some occur under their base form, like *tibtib* ‘vibrate, jew’s harp’ (lit. vibrator); others are derived as locative nouns by *Ca~root-an* reduplication as shown in Table 3, like *ta-tuktuk-an* ‘drum’ (lit. place where to beat) or *ka-kingking-an* ‘bell’.

This sample of ideophones shows that their expressive sensory features are retained even when used descriptively. The inverse correlation between expressiveness and grammatical integration of ideophones thus occurs as a continuum, rather than as bipolar features.

Most ideophones are used as verbs and only about 10% are also used as nouns,<sup>21</sup> in contrast with all other lexical roots which do not show any such categorical restriction in Amis. Instances of ideophones used as nouns by zero derivation are *teki~tekiq* ‘clanking sound’ in (33b), *teraq* ‘a drop’ in example (11c), *badambang* ‘stormy weather, a mess’. They sometimes depict the SOURCE of the sound, for instance, *ciwciw* ‘peep peep’ also means ‘a chick, a duckling’.

IDEOPHONE	DERIVED IDEOPHONIC VERB (DESCRIPTIVE USE)	DERIVED NOUN by <i>Ca~reduplication-an</i> <sup>22</sup>
<i>cikcik</i> ‘sound of chopping’	<i>mi-cikcik</i> ‘slice, chop’	<i>ca-cikcik-an</i> ‘chopping board’
<i>tuktuk</i> ‘sound of hitting’	<i>mi-tuktuk</i> ‘beat’ <i>mi-sa-tuktuk</i> ‘beat the drum’	<i>ta-tuktuk-an</i> ‘drum’
<i>kingking</i> ‘sound of ringing, clink, tinkle’	<i>mi-kingking</i> ‘ring, make a tinkling sound’	<i>ka-kingking-an</i> ‘bell’

Table 3: Some derivational processes of ideophones.

The word *cas*, which depicts a sizzling sound, followed by *sa* as in (34a), also appears under its reduplicated form *cas~cas* to depict the sound of splashing water or the source of that sound, a ‘splattering waterfall’ (34b); in (34c), it is derived as a verb ‘splatter like a waterfall’ and used descriptively.

(34)

- a. *Ca::s sa = tu k-u dangah n-umita.*  
 IDEO.LENGT say = PFV NOM-NM pot GEN-1PL.INCL  
 ‘Sizzle goes the cooking-pot.’ (Katacumuli.041)
- b. *Ma-tengil = aku k-u suni n-u cas~cas.*  
 UV-hear = GEN.3SG NOM-NM sound GEN-NM RDP~IDEO.cascade  
 ‘I can hear the sound of the cascade.’
- c. *Ma-sa-cas~cas k-ina nanum.*  
 NAV-do-RDP~IDEO NOM-NM water  
 ‘The water is splattering like a cascade.’

<sup>21</sup> Common nouns are marked by *u* (NM) and are inflected for case as *k-u* (nominative), *t-u* (oblique), *n-u* (genitive).

<sup>22</sup> *Ca~root-an* reduplication is formed by repeating the root’s initial consonant followed by /a/.

#### 4.2. The semantic extension of ideophones

Ideophones are often polysemous owing (i) to their vague semantics applying by analogy and with context-dependent meaning, or (ii) owing to their metaphorical extension to similar perceptual contexts within close semantic range.

*Telaq* depicts the sound of fire-crackers, rifle shots, applause and thunder claps. *Kelaq* depicts rattling, cracking or creaking sounds, extending to the sound of walking with wooden clogs and to a ‘wooden tipcart’ that makes a rattling sound. *Peri~periq* depicts sounds occurring in fast succession like 1) the sound of a machine-gun, 2) speaking like a machine-gun (35a), by extension 3) the sound of diarrhea, and of 4) a watery mud-slide as in (35b) where the ideophone functions as a modifier.

(35)

- a. *Peri~periq han a s<em>uwal.*  
 RDP~IDEO do.SO LK <UM> speak  
 ‘He spoke like a machine-gun.’
- b. *U ma-peri~periq-ay a lutuk.*  
 NM NAV-RDP~IDEO-MODF LK mountain  
 ‘A mountain-mud-slide.’

*Siw* and *siw~siw* depict a swift motion with a hissing sound. In (36a-b), it is depictive, followed by *sa* ‘say’; in (36b) it depicts an AUDITIVE and VISUAL event, namely the swift motion of a snake; it may also depict the VISUAL event of a shooting-star, only retaining the swift motion. In the last two occurrences, it is derived: as a verb *mi-siwasiw* (with epenthetic /a/) in (36c) with the meaning ‘winnow, chaff’; into the even more abstract meaning ‘select’ (interestingly, English *winnow* has a similar semantic evolution ‘sift in the wind’ and ‘select’) in (36d). In both of these latter examples, it is used descriptively.

(36)

- a. *Siw sa-an.*  
 IDEO say-LV  
 ‘It went whiz.’ (U nipiketun ni Hayan.0059)
- b. *Kalamkam siw~siw sa k-u quner haw.*  
 be.fast RDP~IDEO say NOM-NM snake DISC  
 ‘The snake was fast, it went whiz.’ (U uner a kawas.047)

- c. *Mi-siwasiw t-u tipus.*  
 AV-IDEO.winnow OBL-NM rice  
 ‘(they) winnow the rice.’
- d. *Mi-siwasiw t-u kapah-ay atu raqcus-ay a demak.*  
 AV-IDEO.winnow OBL-NM good-MODF and bad-MODF LK action  
 ‘(they) select the good from the bad actions.’ (U puduc nu pawli.024)

The ideophone *tibtib* ‘pulse, vibrate’ also refers to the ‘jew’s harp/vibrator’, and owing to their similar shapes (namely, a VISUAL feature), it colexifies the ‘shuttle’ used in knitting fishing nets. Note that the non-ideophonic word for the ‘jew’s harp’ is *datuk*.

The two following cases show the semantic extensions of ideophones towards more abstract meanings. *Tera~teraq* depicts the sound of flowing water, the splattering sound of rain or water (see also example (11)). It may also refer to an intense flow of tears in (37a), with vowel lengthening expresses intensity and emphasis.<sup>23</sup> In (37b), it is derived as a deverbal noun and its semantic range extends to a VISUAL experience (blood traces) rather than an auditory one; in this case it is used descriptively.

(37)

- a. *Tera~tera::q sa k-u lusaq.*  
 RDP~IDEO:LENGT say NOM-NM tear  
 ‘His tears dropped and dro::pped.’ (Pangangan.008, 10)
- b. *Mi-dudu t-u ni-ka-teraq-an n-u remes.*  
 AV-follow OBL-NM PFV.NMZ-NFIN-IDEO-NMZ GEN-NM blood  
 ‘(they) followed the blood drippings.’ (u patay ni Calaw Ilikic.048)  
 (Lit. (they) followed the blood that had dripped)

Thus, ideophones with faded or lost sensory depiction favor semantic extension towards more abstract meaning and descriptive use.

## 5. Northern Amis ideophones in a typological perspective

The most frequent ideophones in the corpus depict the sounds of water, wind, animals, human sounds related to bodily functions (‘cough’, ‘breathe’), to speech or

<sup>23</sup> The character is afflicted by constant crying.



cries ('murmur', 'howl'); sounds related to motion and movement ('fall', 'clattering walk'), sounds produced by using tools, weapons ('cut', 'scrape', 'strike', etc.), artefacts and instruments. Some of them depict VISUAL perception associated with fast motion like *siw* referring to a shooting-star, or gestures such as rubbing, brushing off, with or without sound. Some depict MOVEMENTS like falling, tumbling down, shaking, flapping wings. Few of them denote tactile sensations, apart from *kapkap* 'grope' (in the dark). No ideophones depicting shape or texture have been encountered.

Dingemanse (2018: 9) states that cross-linguistically, ideophones generally evoke actions and properties rather than objects. This is generally true of Amis, though some of these ideophones may extend to the entities that produce the sound, or to animals named after their cries.

The semantic domains covered by ideophones in Northern Amis mirror Dingemanse's implicational hierarchy (2012), predicting that if languages have ideophones for semantic domains to the right side of the hierarchy, they must then have ideophones on the left. Sound and movement are predicted to be the most common.

(i) Dingemanse's implicational hierarchy (2012)

SOUND < MOVEMENT < FORM < TEXTURE < OTHER SENSORY PERCEPTIONS

In Amis sound-depicting and sound-&movement depicting ideophones are the most numerous, followed by those associating movement-&visual perception.

(ii) The semantic domains of ideophones in Northern Amis

SOUND < SOUND & MOVEMENT < MOVEMENT & VISUAL PERCEPTION < BAD SMELL

The semantic domains of Japonic ideophones are given in (iii) for comparison, since a fifty-year long Japanese occupation of Taiwan (1895-1945) has left a lasting linguistic impact on Formosan languages, with numerous loan words in present-day Amis.

(iii) The semantic domains of ideophones in Japonic languages (McLean 2021: 528)

SOUND < MOVEMENT < SHAPE < TACTITION < INNER PERCEPTIONS < SMELL < COLOUR < TASTE

In contrast with Japonic ideophones where shape and tactile ideophones stand in the middle of the hierarchy, these are non-existent in Amis. Inner perceptions are also very infrequent in Northern Amis, with the possible exception of *pirpir* 'heart-beat'

and *mirmir* ‘shiver, shake’, with their voiceless oral and voiced nasal consonant alternation. In Japanese, smell and taste are the lowest on the scale; in Northern Amis, bad smells and taste words are defined by specific templates and amount respectively to 15% and 9% of all identified ideophones; however, taste words tend to be de-ideophonized. Further investigation shows that Amis ideophones are most generally not borrowed, with only two traceable loan words, *tera~teraq* ‘sound of rain, of flowing liquid’ possibly from Japanese *tara tara* ‘trickle’, and *kurukuru?* ‘IDEO.tumble.down’ possibly from Japanese *kuru kuru* ‘sound of something rotating or spinning round’. This supports Blust’s (2013: 565) remark that ideophones and onomatopoeia are rarely borrowed in Austronesian languages.

## 6. Ideophones in other Formosan languages

Blust (1988; 2013: 565) argued that in Austronesian languages, many ideophones show regular sound correspondences, like other lexemes. Many of the ideophones in Amis are reflexes of Blust’s reconstructed onomatopoeic PAn roots (see Table 5 in the Appendix); other Austronesian languages, among them Ilocano (Philippines), also have cognates identifiable by regular sound correspondences. Among Formosan languages,<sup>24</sup> a few ideophones are common to Northern Amis, Seediq and Kavalan (ckv; Lee 2017) as shown in Table 4.

NORTHERN AMIS	SEEDIQ	KAVALAN
<i>ʔuhʔuh</i> ‘sound of coughing, cough’	<i>quh</i> ‘sound of coughing’	
<i>tuktuk</i> ‘sound of knocking, beating a drum’	<i>-tuk</i> as a submorpheme of <i>tatuk</i> ‘knock’ (Lee 2017: 206)	<i>tuktuk</i> ‘knocking sound’
<i>rasmus</i> ‘drizzle’ (visual)	<i>ras</i> ‘sound of ‘flowing water’	
<i>telaq</i> ‘sound of rifle shots’	<i>tlelak</i> ‘sound of a machine gun’	
<i>kela~kelaq</i> ‘rattling, cracking, creaking sound; sound of a creaking tipcart’	<i>kerak</i> ‘sound of an old car being driven’	

**Table 4:** Ideophones common to Northern Amis, Seediq and Kavalan (Formosan)

<sup>24</sup> This relies on available published research and needs more investigation.

The ideophones *?ak?ak* in Northern Amis and *ak-ak-ak* in Siraya (fos) depict the cry of ravens and extend to the bird; however, in Siraya, the ideophone is also derived as a verb in (38) extending to mocking, cackling sounds produced by human beings.

(38) Siraya (Formosan; Adelaar 2011: 259)

*Ni-mau-ak-ak-ak ma-tawa tñni-än.*

PAST-AS-cackle-cackle AV-laugh 3sg-LOC

‘They laughed at him with scorn.’ (ix24) (AS = anticipating sequence)

The question of how ideophonic roots are transmitted is commented on by Blust (2009: 357):

“If submorphemic sound-meaning correlations are distributed over a number of genetically related languages in non-cognate morphemes, one must ask how such patterns can be transmitted independently of the forms that exemplify them. There are two logical possibilities: 1) they are transmitted in sets of morphemes which contain a recurrent submorphemic sound-meaning correlation that is then extended to neologisms, or 2) the abstract pattern itself is internalised.”

Much work remains to be done before satisfactory answers can be brought to bear on this question.

## 7. Conclusion

Ideophones are a small subset of the Northern Amis vocabulary (around 4.4% of my 4800 lexical database). They are commonly followed by *sa* ‘say’ or by *han* ‘do so’ and are used to depict sensations (auditory, motion, visual, smells, tastes, etc.) and to convey liveliness and personal reaction. Many ideophones can be used depictively as well as descriptively, in the latter case, they tend to be more grammatically integrated words, i.e., as derived and inflected verbs, more marginally as nouns, yet they retain their original expressive sensory properties.

## Acknowledgements

This research was supported by the LACITO-CNRS and financed by the research strand 3 “Typology and dynamics of linguistic systems” of the Labex EFL (Empirical Foundations of Linguistics) (Investissements d’Avenir, ANR-10-LABX-0083/CGI). It is part of IdEx Paris University (ANR-18-IDEX-0001).

I am very grateful to all Amis consultants and friends for their invaluable collaboration. I also thank the editors of this special issue and two anonymous reviewers for their constructive and thought-provoking remarks that led to some reorganization of this article.

## Abbreviations

1 = 1 <sup>st</sup> person	HORT = hortative	NOM = nominative
2 = 2 <sup>nd</sup> person	IDEO = ideophone	NM = noun marker
3 = 3 <sup>rd</sup> person	IMP = imperative	NMZ = nominalizer
AV = actor voice	INCH = inchoative	OBL = oblique
CA~ = Ca~reduplication	INCL = inclusive	ONOM = onomatopoeia
CAUS = causative	INFX = infix	PFV = perfective
COMP = complementiser	INTENS = intensive	PM = person marker
COS = change of state	LENGT = lengthening	POSS = possessive
CV = conveyance voice	LK = linker	PREP = preposition
DEIC = deictic marker	LOC = locative	PROH = prohibitive
EP = epenthetic	LV = locative voice	PSA = preferred syntactic argument
EXCL = exclusive	MODF = modifier	RDP = reduplication
EXS = existential	NAV = non-actor voice	
FUT = future	NEG = negation	
GEN = genitive	NFIN = non-finite	

## References

- Adelaar, Alexander K. 2011. *Siraya Retrieving the Phonology, Grammar and Lexicon of a Dormant Formosan Language*. Trends in Linguistics Documentation 30. Berlin/Boston: De Gruyter Mouton.
- Akita, Kimi. 2017. The linguistic integration of Japanese ideophones and its typological implications. *Canadian Journal of Linguistics*, 62(2). 314–333.

- Ameka, Felix K. 2001. Ideophones and the Nature of the Adjective Word Class in Ewe. In Voeltz, F. K. Erhard & Kilian-Hatz, Christa (eds.), *Ideophones*, 25–48. Amsterdam: John Benjamins.
- Bergen, Benjamin K. 2004. The psychological reality of phonaesthemes. *Language* 80. 290–311.
- Blust, Robert. 1988. *Austronesian Root Theory; an essay on the limits of morphology*. SLCS 19. Amsterdam: John Benjamins.
- Blust, Robert. 2003. The phonestheme  $\eta$ - in Austronesian languages. *Oceanic Linguistics* 42. 187–212.
- Blust, Robert. 2009. *The Austronesian languages*. Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.
- Blust, Robert. 2013. *The Austronesian languages*. (Revised version). Asia-Pacific Linguistics; Canberra, Australian National University.
- Blust, Robert. 2022. *A Dictionary of Austronesian Monosyllabic Roots (Submorphemes)*. Berlin: Walter de Gruyter.
- Blust, Robert & Stephen Trussel. *The Austronesian Comparative Dictionary (ongoing)*. Available online at: <http://ling.ill.hawaii.edu/dicts/ACD/>.
- Bril, Isabelle. 2016. Information Structure in Northern Amis (Formosan): a morphosyntactic analysis. *Oceanic linguistics*, 55(2). 451-481.
- Bril, Isabelle. 2017. Roots and stems: Lexical and functional flexibility in Amis and Nêlêmwa. *Studies in Language* 41(2). 358-407.
- Bril, Isabelle & Skopeteas, Stavros. 2021. The syntax and prosody of focus in Northern Amis (Formosan). In Haude, Katharina, Palancar, Enrique & Vanhove, Martine (eds.). *Special issue on Clefts. Faits de Langue*. 51(2). 61-87.
- Bril, Isabelle. 2022. Lexical restrictions on grammatical relations in voice constructions (Northern Amis). In Eva van Lier and Maria Messerschmidt (eds.), *Lexical restrictions on grammatical relations in voice and valency constructions*. [Special issue]. *STUF Language Typology and Universals*, 75(1). 21-71.
- Chen, Teresa M. 1969. *Natauran texts*. Unpublished manuscript.
- Dingemanse, Mark. 2011. *The meaning and use of ideophones in Siwu*. Doctoral Dissertation, Radboud University, Nijmegen.
- Dingemanse, Mark. 2012. Advances in the Cross-Linguistic Study of Ideophones. *Language and Linguistic Compass* 6(10). 654–672.
- Dingemanse, Mark. 2017. Expressiveness and system integration: On the typology of ideophones, with special reference to Siwu. *STUF* 70(2). 363–384.

- Dingemanse, Mark. 2018. Redrawing the margins of language: Lessons from research on ideophones. *Glossa: a journal of general linguistics* 31(4). 1–30.
- Dingemanse, Mark. 2023. Ideophones. *The Oxford Handbook of Word Classes*, ed. By Eva van Lier. Oxford: Oxford University Press. Preprint doi: 10.31234/osf.io/u96zt.
- Dingemanse, Mark & Akita, Kimi. 2017. An inverse relation between expressiveness and grammatical integration: On the morphosyntactic typology of ideophones, with special reference to Japanese. *Journal of Linguistics*, 53(3). 501–532.
- Fey, Virginia. 1986. *Amis dictionary*. Taipei: Evangelical Alliance Mission.
- Ibarretxe-Antuñano, Iraide. 2017. Basque ideophones from a typological perspective. *The Canadian Journal of Linguistics/La revue canadienne de linguistique* 62(2). 196–220.
- Klamer, Marian. 2001. Expressives and iconicity in the lexicon. In Voeltz, F.K. Erhard & Kilian-Hatz, Christa (eds.), 165-182.
- Lee, Amy Pei-jung. 2015. Lexical categories and conceptualization of olfaction in Amis. *Language and Cognition*, 7(3). 321-350.
- Lee, Amy Pei-jung. 2017. Ideophones, Interjections, and Sound Symbolism in Seediq. *Oceanic Linguistics*, 56(1). 181-209.
- Li, Paul Jen-kuei & Tsuchida, Shigeru. 2009. Yet more Proto Austronesian infixes. *Discovering history through language: papers in honour of Malcolm Ross*, 345–362. Canberra: Pacific Linguistics.
- McLean, Bonnie. 2021. Revising an implicational hierarchy for the meanings of ideophones, with special reference to Japonic. *Linguistic Typology*, 25(3). 507–549. <https://doi.org/10.1515/lingty-2020-2063>
- Poinsot, Maurice & Pourrias, Louis. 1966. *Dictionnaire amis-français*. Paris: Société des Missions Etrangères de Paris.
- Rata, Namoh. 2013. *Ameizuyu cidian [Amis dictionary]*. Taipei: Nantian shuju [SMC Publishing Inc.].
- Reid, Lawrence. 1994. Morphological Evidence for Austric. *Oceanic Linguistics*, 33(2). 323-344.
- Rubino, Carl. 2001. Iconic morphology and word formation in Ilocano. In Voeltz, F.K. Erhard & Kilian-Hatz, Christa (eds.), 303-320.
- Teng, Stacy Fang-Ching. 2008. *A Reference Grammar of Puyuma*. Canberra: Pacific Linguistics 595.
- Voeltz, F.K. Erhard & Kilian-Hatz, Christa (eds.). 2001. *Ideophones*. TSL 44. Amsterdam: Benjamins.

Wu, Joy Jing-lan. 2006. *Verb classification, case marking, and grammatical relations in Amis*. PhD dissertation. New York, NY: University at Buffalo, the State University of New York.

Zeitoun, Elizabeth. 2023. Ideophones vs. onomatopoeia in Tungho Saisiyat. Paper read at the Bi-monthly Colloquium of the Institute of Linguistics at Academia Sinica. March 13, 2023. Taipei: Academia Sinica.

**CONTACT**

isabelle.bril@cnrs.fr

## Appendix

Table 5 lists some ideophones in Northern Amis, similar words in a few other Formosan languages, with their PAN cognates.

NORTHERN AMIS	other FORMOSAN LANGUAGES	PAn (Blust 1988)
'ak'ak 'to/a crow, raven'	Puyuma 'ak'ak '(a) crow' Siraya ak-ak-ak '(a) crow'	*kak 'cackle' (1988: 101)
bek 'thud'		*bek 'pound, thud' (1988: 27, 38)
bengbeng 'bang'		*beng 'dull resounding sound'
bikbik 'flutter'	Puyuma bikbik 'shake off' (Teng 2008: 36)	
pikpik 'flap wings'		*pik 'click' *pikpik 'sound of patting, tapping'
kapkap 'grope'		*kapkap 'feel in the dark, grope' (1988: 103)
k(e)ris 'scratch'		*ris 'scratch a line' (1988: 147)
kingking 'ring, tinkle'		*kiŋ, *kiŋkiŋ <sub>1</sub> 'ringing sound' (1988: 41)
kung 'sound of a big bell' kungkung 'knock with hollow thump', 'beat a drum'	Seediq kung 'sound of a door being knocked'	*kungkung 'deep resounding sound', 'a slit-gong, a hollowed tree trunk used to send messages'
kiskis 'scrape, scratch' kihkih 'scrape, scratch'		*kis 'scratch' PMP *kihkih (Blust 2013: 643)
kutkut 'scrape, scratch'		*kutkut <sub>2</sub> , kudkud 'scrape with metal' (Blust & Trussel)
pekpek 'pound, hammer'		*pek 'clap, slap, thump' *pekpek 'beat, hit'
pukpuk 'beat & husk rice' (in a wooden tank)		*puk 'clap, slap, thump' (1988: 38) *pukpuk 'hammer, pound, beat' (Blust & Trussel)
ritrit 'cut, reap' (grain)		*riC 'sound of ripping' (1988: 80)
tekiq 'clank'		*tekiq 'sound of rock hitting'
tektek 'hit a post in the ground'		*Tek 'light knocking sound' (1988: 81)



NORTHERN AMIS	other FORMOSAN LANGUAGES	PAn (Blust 1988)
<i>tiktik</i> ‘pat, tap’	Puyuma <i>tiktik</i> ‘hammer at’ (Teng 2008: 36)	*Tik ‘crack, click, tick, tock’ (1988: 27, 38)
<i>tuktuk</i> ‘sound of hitting’	Kavalan <i>tuktuk</i> ‘knocking sound’ Seediq <i>-tuk</i>	*-tuk ‘knock, pound, beat’ *tuktuk ‘hit with a hammer’

**Table 5:** Ideophones and their PAn cognates.

CaŋCVC roots (North Amis)	PAn *(C)an(e)CV(C) (Blust 1988: 60)	banCVC roots (North Amis)	PAn *banCVC (Blust 1988: 60)
?anglis ‘smell of fish’	*qaŋeRiS ; *qaŋ(e)seR ‘stink as of urine’ *qaŋeSit ‘stink of a skunk; smell of some plants’	<i>bangsit</i> ‘stench of urine’	*ban(e)sit ‘stench’

**Table 6:** Smell ideophones in Amis and their PAn cognates.

# Ideophonic verb compounds in Archi

GILLES AUTHIER

ÉCOLE PRATIQUE DES HAUTES ÉTUDES – PSL

Submitted: 21/03/2023 Revised version: 07/09/2023

Accepted: 12/02/2024 Published: 12/06/2024



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

## Abstract

Archi is a one-village language (1,000 speakers) belonging to the Lezgian branch of East Caucasian. It has a large and productive class of compound verbs combining a light verb and ‘coverbs’ of nominal, adjectival, verbal or unknown origin. It stands out among its closest relatives in the way it has created, under the influence of Lak, a large class of compounds using the verb *bos* ‘say’. We consider all coverbs used with *bos* and its allomorphs to be ideophones and propose a semantic classification of all ideophonic verbal compounds: sound and speech verbs, verbs of non-auditory sensations, ingestion, movement, and effortful activities. All primary data were extracted with partial paradigms and some examples from Chumakina et al.’s Archi online dictionary. Their phonotactic shapes, not substantially different from other parts of speech, are examined, as well as borrowings and specific ‘children’s speech’ ideophones.

**Keywords:** East Caucasian; Archi; compound verbs; ideophones

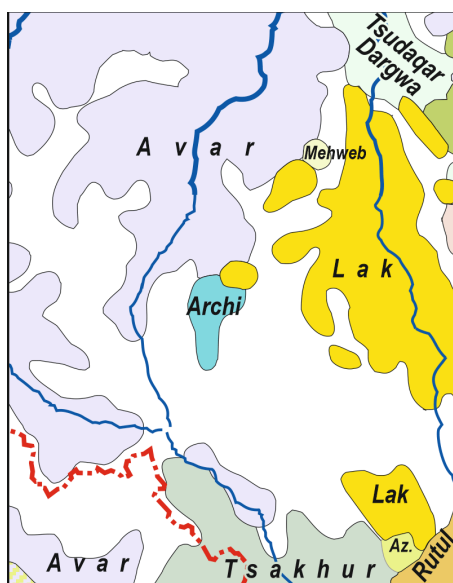
## 1. Introduction

While all languages have symbolic uses of certain combinations of phonemes to suggest sounds or other sensations perceived by speakers, it is clear that not all languages use them with the same frequency. In some parts of the world, languages are known to make constant use of a variety of ideophones, while in other regions ideophones are much more discreet, or even difficult to find for descriptive linguists. More surprisingly, it happens that within the same language family, spoken in the

same region and among cultures that are closely related, some languages use very few ideophones, while in others they are omnipresent. This is the case in the East Caucasian family, where most languages make only very limited use of ideophones. However, three geographically contiguous languages, two large ones (Dargwa and Lak) and one very small (Archi), belonging to three different branches of the family, have many verbal compounds, extremely frequent in all types of speech, which consist of the verb ‘say’ immediately preceded by and often merged with a ‘coverb’ whose expressive function is often but not always clear.

With about 1,000 speakers in total, Archi has also been one of the most studied and best described languages in this family, especially since Kibrik et al. (1977), Kibrik (1994) and Chumakina et al. (2007). It has also gained some fame for its rich verbal morphology, which Kibrik estimates to produce up to 1.5 million inflected and derived forms. Classified as a member of the Lezgian subfamily, Archi has no contact with its closely related cousins, from which it has been separated to the south by the Great Caucasus mountain range for at least one thousand years. This geographical separation positions Archi as a distinct northern outlier.

On the other hand, Archi has long been in contact with Lak in the northeast and with Avar in the west. Today the speakers identify as ethnic Avars, although Russian has replaced Avar as the language of instruction in the village school. Nevertheless, the use of ideophones in ‘say’-compounds bears witness to the earlier predominance of Lak, rather than Avar, on Archi. Figure 1 shows a map of Archi and surrounding languages.



**Figure 1:** Location of Archi and surrounding languages (courtesy of Y. Koryakov)

### 1.1. Simplex verbs in East Caucasian, Lezgian and Archi

Within the East Caucasian family, western branches (Nakh, Avar, Andic, and Tsezic) have overall simpler verbal morphology than eastern branches (Dargwa, Lezgian and Khinalug), with Lak occupying an intermediate position on the complexity scale. This state of affairs correlates largely with a higher number of simplex verbs and a lower number of compounding strategies in the western branches as opposed to the inflectionally more complex eastern ones.

Within the Lezgian subgroup, some languages have many simplex verbs (several hundred in Rutul, Tabasaran or Budugh, each of which belongs to one of the three branches of the so-called Central Lezgian languages), while others have fewer than two hundred, such as Kryz (sister language of Budugh), Tsakhur (sister language of Rutul) or Udi (another outlier to the south), although this does not seem to be linked to their immediate relationship.

Archi is a northern outlier, and belongs to the second category, with only 163 simplex verbs according to Kibrik et al. (1977) via Chumakina and Corbett (2015: 102). These simplex verbs typically mark at least aspect and gender, as in (ex. 1; note that the ‘animate’ gender includes many inanimate referents, like ‘corner’ here, or ‘road’ in ex. 2a):

- (1) *to-w-mu*                      *χ:<sup>w</sup>alli-li-n*                      *mec:’e*                      *b-eq:’u.*  
 DIST-M-ERG                      bread-OBL-GEN      corner(A)(NOM)      A-bite.PF  
 ‘He bit off a piece of bread.’ (Chumakina et al. 2007, as all further examples unless otherwise stated)

As in other Lezgian languages preserving the category of gender and gender/number agreement of verbs with the nominative argument (S or P), the morphology of Archi verb stems is complex, characterized by the irregular interplay of gender/number markers with aspect markers, with numerous verb classes. Archi also preserves a few traces of spatial prefixes (preverbs). In the paradigm of such verbs, when two forms appear in one cell, the former represents the fourth gender (N, for ‘non-animate’) form while the latter represents the third gender (A for ‘non-human animate’) form. For instance, ‘do’, ‘bite’, ‘become’ and ‘beat’ have the following principal parts (Table 1):

	INF	IPF	PF	IMPER	MSD
‘do’	<i>as/ ábas</i>	<i>ar/ bar</i>	<i>aw/ ábu</i>	<i>a/ ba</i>	<i>ámul/ bámul</i>
‘become’	<i>kes/ bekés</i>	<i>ker/ bekér</i>	<i>ét:i/ ébt:i</i>	<i>ka/ baká</i>	<i>kummúl/ bukmúl</i>
‘beat’	<i>dáχis/ dábyχis</i>	<i>dárχir/ dábyχir</i>	<i>dáχdi/ dábyχdi</i>	<i>dáχi/ dábyχi</i>	<i>dáχmul/ dábyχmul</i>
‘bite’	<i>éq:’as/ béq:’as</i>	<i>érq:’ur/ bérq:’ur</i>	<i>réq:’u/ béq:’u</i>	<i>éq:’a/ béq:’a</i>	<i>éq:’mul/ ébq:’mul</i>

**Table 1:** Inflection of simplex verb stems in the third and fourth gender in Archi  
(data from Chumakina et al. 2007)

If we consider the situation in languages such as Budugh, Rutul and Tabasaran, it does not seem that the maintenance of a large number of simplex verbs correlates with simplification of the inflectional system, and therefore the reduction of the number of simplex verbs in Archi is not necessarily linked to the maintenance (or even the development) of a rich and complex verbal morphology.

Instead, contact with two major prestige languages, which are only distantly related to Archi, has undoubtedly favoured the renewal of its verbal lexicon, in the same way that Azeri has contributed to the replacement of many verbs in many Lezgian languages. Contact with Avar has even affected verbal inflection: for instance, the Avar prohibitive suffix *-gi* has replaced in Archi the equivalent prefix *ma-* attested in all other Lezgian languages.

### 1.2. Compound verbs in East Caucasian, Lezgian and Archi

Whatever the cause of the reduced number of simplex verbs in Archi, the language makes extensive use of compound verbs to compensate for this reduction. As in Kryz, Udi or Tsakhur, there are many more or less lexicalized and morphologically merged locutions in Archi combining the verbs *as* ‘do’ (ex. 2a and 2b), *kes* ‘become’ (ex. 3) with adjectives or ‘stative verbs’ (themselves a type of adjectives):

- (2) a. *χaballi de<sup>h</sup>q’ kut’a b-ar.*  
 talk.ERG road(A) short A-do.IPF  
 ‘Conversation shortens the journey.’

- b. *χ:allu-r t:anna bošor χala w-ar.*  
 bad.ATTR-F woman.ERG man(NOM) old M-do.IPF  
 ‘A bad wife makes the husband old.’

- c. *ez ja-t kummul aχ<sup>s</sup>*  
 (N)1SG.DAT PROX-N food(N)(NOM) (N)enough  
*ker.*  
 (N)become.IPF  
 ‘This (amount of) food is enough for me.’

This type of compound is known to all languages related to Archi, whether closely or distantly, but with a level of productivity that varies greatly across these languages, and is clearly inversely proportional to the number of simplex verbs they possess.

Archi differs notably from all other Lezgetic languages in its use of verbal compounds with the verb ‘say’. This strategy, unknown in Avar, is widespread in Lak, as well as in all Dargwa dialects. Given the long-dominant position of Lak in the region and in spite of Dargwa speakers now outnumbering Lak speakers four to one, it is fairly certain that the development of ‘say’-compounds spread from Lak into the Dargwa dialects in the northeast and into Archi in the southwest.

Note that some Archi verbs corresponding to verbs often involving ideophones in other languages may not fall into this category. For instance, the verb meaning ‘laugh’ bears no trace of any light verb, and is indeed related, as seen in Table 2, to simplex verbs in other Lezgetic languages. It was probably never a compound verb, and unlike all compound verbs formed with ‘say’, it takes gender/number prefixes. This, of course, does not rule out the possibility that the root ( $\chi^s a$ ) of this verb was originally an ideophone (cf. Russian *xoxotat*’, etc).

	INF	IPF	PF	IMPER	MSD
‘laugh’	$\chi^s arás$ /	$\chi^s ar$ /	$\chi^s ará\chi^s u$ /	$\chi^s ará\chi^s a$ /	$\chi^s ará\chi^s -mul$ /
	$ba\chi^s rás$	$ba\chi^s ár$	$ba\chi^s rá\chi^s u$	$ba\chi^s rá\chi^s a$	$ba\chi^s rá\chi^s -mul$

**Table 2:** Principal parts of the verb ‘laugh’ in Archi  
 (data from Chumakina et al. 2007)

### 1.3. The verb ‘say’ in Archi, forms and external comparison

The verb ‘say’ in Archi is simple, with four out of five principal parts showing the onset *b*, while the imperfective (present tense) form starts with the consonant *w* (cf. Table 3).

	INF	IPF	PF	IMPER	MSD
'say'	<i>bos</i>	<i>war</i>	<i>bo</i>	<i>ba</i>	<i>bumul</i>

**Table 3:** Principal parts of the verb 'say' in Archi

This is a case of quasi-suppletion, but the *b/w* onset is probably a frozen animate gender marker, found with speech acts (these belong to the non-human animate gender in languages like Rutul or Kryz, cf. Authier 2009§ 4.3.1).

This verb has reflexes in all other Lezgian languages:

- Southern branch: Budugh perfective *yɪpa*, imperfective *yuʔu*, Kryz masdar (verbal noun) *lip-idž*, imperfective *lu-* or *li-*;
- Eastern branch: Lezgian masdar *luhu-n*, perfective *laha-*, Tabasaran masdar *pub*, Aghul infinitive *pas*;
- Western branch: Rutul infinitive *(hu)xus*, imperfective *ruxu-* or *rux<sup>w</sup>a-*, Tsakhur infinitive *ehes*, perfective *ivho*;
- Southern outlier: Udi masdar *pe-sun*.

These cognates allow one to reconstruct a root *\*h<sup>w</sup>* or *\*x<sup>w</sup>*. Anlauts with *l-* in Lezgian and Kryz probably feature a preverb meaning 'on' because this verb was used to express the meaning 'bestow a name'. Those with a labial stop retain the frozen non-human animate gender prefix *\*b-* or *v* in Tsakhur. Agreement with an animate speech act is linked on this root with perfectivity. Imperfective forms lack it, and another, imperfective root *k* 'tell' is attested in all three core branches (in Tsakhur, in Budugh, as well as in Tabasaran, where it acts as a suppletive root for imperfective tenses of 'say').

#### 1.4. Morphosyntax of the verb 'say' in Archi

The verb 'say' is a ditransitive verb: its Agent is marked with the ergative case, while its Patient (the speech content) can be a noun or a substantivized modifier in the unmarked nominative case, and the addressee takes the Archi contlative<sup>1</sup> case, as in (ex. 3):

<sup>1</sup> Many East Caucasian languages have a series of spatial cases denoting 'contact with a vertical landmark', cf. Ganenkov (2005).

- (3) *zari wa:r-šī os č'at bo-qi.*  
 1.ERG 2.SG.OBL-CONTLAT one word(N) say.PF-FUT  
 'I will say a word to you.'

The class of ditransitive verbs is very large in most Lezgian languages and more broadly in East Caucasian because it contains the large subclass of contact verbs which treat body parts or instruments as the unmarked argument, in the nominative case (ex. 4 and 5), and even if these referents are omitted, the third (goal) argument, if animate, retains its oblique marking (ex. 6).

- (4) *zari to-w-mu-s χ:ik' dabχdi.*  
 1SG.ERG DIST-M-OBL-DAT fist(A)(NOM) A.hit.PF  
 'I hit him with my fist.'

- (5) *zari to-w-mu-s χatum-l-a-k č'ele dabχdi.*  
 1SG.ERG DIST-M-OBL-DAT chest-OBL-IN-LAT stone(A)(NOM) A.hit.PF  
 'I hit him in the chest with a stone.'

- (6) *bošor-mu łanna-s darχir-šī i.*  
 man-ERG woman.OBL-DAT hit.IPF-CV (N)COP  
 'The husband is beating his wife.'

The verb 'say' is used in a somewhat similar manner, and many speech act expressions registered in the dictionary, including combinations like these proverbs (ex. 7 and 8):

- (7) *e'nkdu-mmu-s q'e'jla akbar war-t'u.*  
 deaf-OBL-DAT two\_times azan(N) say.IPF-NEG  
 'For a deaf person, one does not issue a call for prayer twice.'

- (8) *hara-χut mu-ttu-t war-ttu-mu muχara-χut*  
 front-TRANS NICE-ATTR-N say.IPF-ATTR-ERG behind-TRANS  
*χ:allu-t war.*  
 bad.ATTR-N say.IPF  
 'He who says good things to his face says bad things behind his back.'



Other lexicalized expressions, such as *k<sup>w</sup>ač'et bos* 'lie', are equally not ideophonic because they contain a nominal element which exists independently (the word *k<sup>w</sup>ač'et* is a noun in the nominative case; it has a plural *k<sup>w</sup>ač'et-mul* and other cases such as ergative singular *k<sup>w</sup>ač'et-li* and ergative plural *k<sup>w</sup>ač'et-mul-čaj*) and holds the status of syntactic constituent, since the unmarked nominative case is proper to any patient argument of a transitive verb. The combinations of such nouns or noun phrases with the verb *bos* are thus locutions, not compounds, as illustrated in Table 4.

	PF	IPF	INF	IMPER	MSD
'lie'	<i>k<sup>w</sup>ač'et bos</i>	<i>k<sup>w</sup>ač'et war</i>	<i>k<sup>w</sup>ač'et bo</i>	<i>k<sup>w</sup>ač'et ba</i>	<i>k<sup>w</sup>ač'et bumul</i>
'greet'	<i>warč'ámi bos</i>	<i>warč'ámi war</i>	<i>warč'ámi bo</i>	<i>warč'ámi ba</i>	<i>warč'ámi bumul</i>
'sing the azan'	<i>akbár bos</i>	<i>akbár war</i>	<i>akbár bo</i>	<i>akbár ba</i>	<i>akbár bumul</i>
'thank'	<i>bárka bos</i>	<i>bárka war</i>	<i>bárka bo</i>	<i>bárka ba</i>	<i>bárka bumul</i>
'tell the truth'	<i>t'államat bos</i>	<i>t'államat war</i>	<i>t'államat bo</i>	<i>t'államat ba</i>	<i>t'államat bumul</i>

Table 4: Non-ideophonic compounds with the verb 'say' in Archi  
(data from Chumakina et al. 2007)

### 1.5. Imperative words as coverbs

Archi also has "imperative interjections" like *ma!* 'Take this!' or *č'eb(a)!* 'come!' (ex. 9 and 10) which can be employed with the verb 'say':

- (9) *lač<sup>w</sup>ejwu*                      *to-w-u*                      *č'eb + ba!*  
 1PL.INCL.COM.M                      DIST-M-AND                      "come!" + say.IMPER  
 'Ask him to come with us.'

- (10) *w-ez*                      *to-w*                      *za-łu*                      *kino-li-tik*  
 M-1SG.DAT                      DIST-M                      1.OBL-COM                      cinema-OBL-SUPEREL  
*č'eb + bo-s*                      *kł'an-ši*                      *w-i*.  
 "come!" + say-INF                      want-CV                      M-COP  
 'I want to ask him to come to the cinema with me.'

The principal parts of these combinations sometimes show no trace of morphological modification of the light verb. Nevertheless, the proper imperative

verb forms like *(b-)oq'i* ‘Give it to me!’ (a suppletive imperative form of the verb *kłos* ‘give’), used with ‘say’ as a compound verb meaning ‘request’, can be followed by either an unchanged form of the verb ‘say’ or, mainly in the imperfective, by a reduced form. Such a reduced form has grammaticalized as a ‘reportative’<sup>2</sup> ending *-er*, which goes back to *war*, the full imperfective verb stem of *bos* ‘say’. Compare the direct and indirect constructions in (ex. 11a and 11b):

(11) a. *os q<sup>ws</sup>et'u titrat oq'i-su!*  
 one two.N notebook(N) give.IMPER-PLEASE  
 ‘Please give (me) one or two copybooks.’

b. *jo-w-mu žu-s jamu-t surat*  
 PROX-M.-ERG REFL.M.OBL-DAT PROX-N.SG picture(N)  
*oq'-er.*  
 give\_me.IMPER-RPRT  
 ‘‘Give me this photo,’’ he says.’

Table 5 shows some such locutions based on an imperative form:

	PF	IPF	INF	IMPER	MSD
‘offer’	<i>má bos</i>	<i>má war</i>	<i>má bo</i>	<i>má ba</i>	<i>má bumul</i>
‘invite’	<i>č'éba bos</i>	<i>č'ébuwar</i>	<i>č'éba bo</i>	<i>č'éb(a) ba</i>	<i>č'éba bumul</i>
‘request’	<i>óq'i bos</i> / <i>b-óq'i bos</i>	<i>óq'i war</i> / <i>óq'er</i> <i>bóq'i war</i> / <i>b-óq'er</i>	<i>óq'i bo</i> / <i>b-óq'i bo</i>	<i>óq'i ba</i> / <i>b-óq'i ba</i>	<i>óq'i bumul</i> / <i>b-óq'i bumul</i>

Table 5: Non-ideophonic compounds with frozen imperative words  
 (data from Chumakina et al. 2007)

### 1.6. Proper and pseudo-ideophonic coverbs

Ideophonic coverbs have been identified but not properly studied in the existing literature on Archi. They are coalescent words, never found without the verb ‘say’ (i.e. they do not constitute a lexical entry in their own right), with which they often

<sup>2</sup> On this reportative suffix, see Chumakina (2013).

fuse, especially with the imperfective form *war*: ‘yell’ *q’akar* < *q’ik* + *war* and ‘cry’ (baby) *a<sup>s</sup>ngá<sup>r</sup>* < *a<sup>s</sup>ngá<sup>s</sup>* + *war*.

The root of the verb ‘say’ can take the form *bu-* or *mu-*, instead of *bo-*, as with the verb ‘cry’ in (ex. 12):

- (12) *e<sup>s</sup>mmu-li*            *ja-r*    *lo*            *q’ik* + *bo-li*.  
 cry + say.PF-EVID    PROX-F    child(NOM)    yell + say.PF-EVID  
 ‘The girl cried and screamed.’

In the perfective form *e<sup>s</sup>mmuli* (< \* *e<sup>s</sup>n* + *bo-li*) both consonants at the morpheme break have assimilated. However, the coverb *e<sup>s</sup>n* is as identifiable in *e<sup>s</sup>mmuli* as *qi<sup>r</sup>* is in *qi<sup>r</sup>-boli*. In both cases, the elements *e<sup>s</sup>n*, *qi<sup>r</sup>* and others of the same type are likely to be ideophones from the outset. Compare in (ex. 13) the analytic verb forms of ‘cry’ with the synthetic forms of ‘laugh’:

- (13) *e<sup>s</sup>n*    *war-ma*            *χ<sup>w</sup>a<sup>r</sup>-gi*,            *χa<sup>r</sup>-ma*  
 cry    say.IPF-WHERE    laugh.IPF-PROHIB    laugh.IPF-WHERE  
*e<sup>s</sup>n*    *war-gi!*  
 cry    say.IPF-PROHIB  
 ‘Where one cries, do not laugh! Where one laughs, do not cry!’

Among the verbs listed in the following sections, some form their masdar with the default masdar suffix *-mul*, which may be related to *-mu-s* < *-bus* < *bos*, while others instead use a suffix *-t’i*, which is never found attached to roots of simplex verbs. This suffix is actually borrowed from Lak, where it is the masdar of the verb *t’un* ‘say’. It is a curious paradox that the masdar form of many compound verbs otherwise formed with a variant of *bos* shows no trace of the Archi verb *bos*.

Most of these verbs have semantics compatible with an origin as ideophones, encompassing a range of semantic subdomains, from intransitive sound verbs to monotransitive action verbs. Among these supposed ideophonic compounds of Archi, Chumakina (2015) identifies two main semantic classes: ‘verbs of speech and mouth activities’ and ‘ideophones’, adding: ‘However not all *-bos* verbs belong to these semantic classes, [and] the rest of them do not constitute a homogenous class.’ It seems necessary to continue the examination from a semantic perspective, before drawing conclusions as to the more formal aspects, in order to arrive at a more detailed description of this essential part of the Archi verbal lexicon.

The semantic fields to which coverbs used with ‘say’ belong correspond to those generally illustrated by ideophones in the other languages of the world where they are used. Whatever their origin, they cannot be used without the light verb ‘say’.

In the infinitive of such verb compounds, the light verb may appear in three somewhat integrated forms: *bos*, *-bus* and *-mus*, e.g.

- *hakɫ’ bos* ‘yawn’
- *c’á-bus* ‘drink cold liquids’,
- *ʕám-mus* ‘bite’ (child speech)

In what follows, we shall first consider as ideophones all coverbs used with *bos* or its bound variants (*-bus* and *-mus*) because in a language as phonologically rich as Archi, which contains no fewer than 70 consonants, it would be hazardous to decide a priori which phonemic combinations should be considered expressive. It shall therefore be assumed that all non-autonomous coverbs used with a form of the verb ‘say’ (*bos*, *-bus* or *-mus*) are essentially ideophonic.

The ideophonic compound verbs will be presented in turn and classified in the following semantic fields: first, sound verbs (oral, body-related, non-body-related) will be discussed in Section 2, followed by non-sound verbs (non-auditive sensations, movements, actions) in Section 3. Section 4 will focus on phonological issues, Section 5 will deal with derivation, Section 6 will present ideophones in children speech, Section 7 will address the treatment of loans, and finally, Section 8 will demonstrate, through external comparison, that some supposed ideophonic coverbs in fact have a non-ideophonic origin.

## 2. Sound verbs

This section will be devoted to verbs denoting the production of sound. It is divided into two subsections, the first dealing with sounds of the external, natural world, and the second all nuances of oral human sounds, i.e. speech verbs.

### 2.1. Weather and animal sound verbs

Weather and animal sound verbs are shown in Table 6; they comprise verbs for the sound of rain and thunder (ex. 14), the cries of wild and domestic animals (ex. 15 and following), as well as human cries addressing domestic animals (ex. 22). All these verbs

are intransitive, and all preserve intact the form of the verb ‘say’ in the infinitive (*bos*), except for two where it changes to *-bus*. However, the imperfective forms can be fused.

(14) *dunil qubu.*

sky(N) thunder + say.PF

‘It thundered.’

(15) *ak:ommis heleku uriʔiʔuʔ war.*

dawn.DAT rooster(A) crow say.IPF

‘The rooster crows in the morning.’

(16) *ziz bo-nnu-b tʔantʔ-li imcʔ a-r-tʔu.*

buzz say.PF-ATTR-A bee-ERG honey(N) (N)do-IPF-NEG

‘Not every bee that buzzes makes honey.’ (proverb)

(17) *noʔʂ hihir-ši b-i.*

horse(A) whinny.say.IPF-CV A-COP

‘The horse is whinnying.’

(18) *χʔurkʔ bur-ši b-i.*

pigeon(A) coo.say.IPF-CV A-COP

‘A pigeon is cooing.’

(19) *χʔon bu bo.*

cow(A) moo say.PF

‘A cow mooed.’

(20) *gʷači iqan haʔmpar-ši b-i.*

dog(A) all\_day bark.say.IPF-CV A-COP

‘The dog barks all day.’

(21) *noʔʂ uh war-ši b-i.*

horse(A) sigh say.IPF-CV A-COP

‘The horse is sighing.’

- (22) *uł-mi ju bo č'abu.*  
 shepherd-ERG ideoph say.PF sheep.PL.NOM  
 'The shepherd shouted to drive the sheep on.'

Verbs denoting animal sounds can be used metaphorically to describe a human being (ex. 23–25) or even an inanimate referent (ex. 26):

- (23) *to-w dogi bana haʔar-ši w-i.*  
 DIST-M donkey(A) LIKE bray.say.IPF-CV M.-COP  
 'He brays like a donkey.'

- (24) *te-b χom č'irq'ir war-ši.*  
 DIST-PL.NOM woman.PL.NOM chirp say.IPF-CV  
 'Those women chirp rather than talk.'

- (25) *to-w noq:ʔon bana c'arar-ši w-i.*  
 DIST-M(NOM) mouse(A) LIKE squeak.say.IPF-CV M-COP  
 'He is squeaking like a mouse.'

- (26) *dakł' c'arar-ši i.*  
 door(N) creak.say. IPF-CV (N)COP  
 'The door is creaking.'

	PF	IPF	INF	IMPER	MSD
'thunder'	<i>qúbus</i>	<i>qúr</i>	<i>qúbu</i>	<i>qú ba</i>	<i>qút'i</i>
'pour' (rain)	<i>wéʒ bos</i>	<i>wéʒ war</i>	<i>wéʒ bo</i>	<i>wéʒ ba</i>	?
'chirp'	<i>č'irq'ir bos</i>	<i>č'irq'ir war</i>	<i>č'irq'ir bo</i>	<i>č'irq'ir ba</i>	<i>č'irq'irt'i</i>
'buzz'	<i>zíz bos</i>	<i>zíz war</i>	<i>zíz bo</i>	<i>zíz ba</i>	<i>zíz bumul</i>
'bark'	<i>hámp bos</i>	<i>hámpár</i>	<i>hámp bo</i>	<i>hámp ba</i>	<i>hámp't'i</i>
'whinny'	<i>hihí bos</i>	<i>hihi wár</i>	<i>hihí bo</i>	<i>hihí ba</i>	<i>hihi bumúl</i>
'bray'	<i>hóʔ bos</i>	<i>háʔár</i>	<i>hóʔ bo</i>	<i>hóʔ ba</i>	<i>hóʔ't'i</i>
urge sheep on'	<i>jú bos</i>	<i>júwar</i>	<i>jú bo</i>	<i>jú ba</i>	<i>jú bumul</i>

**Table 6:** Sample of ideophonic weather sound verbs and animal communication verbs in Archi (data from Chumakina et al. 2007)

## 2.2. Speech impairments and inarticulate oral expression

Equally formed by compounding are intransitive verbs denoting speech impairments as well as inarticulate human oral expression (cf. the verbs shown in Table 7, ex. 27–29):

	PF	IPF	INF	IMPER	MSD
‘stammer’	ʔáls bos	ʔáls war	ʔáls bo	ʔáls ba	ʔálst'i
‘grumble’	dúmɛum bos	dúmɛum war	dúmɛum bo	dúmɛum ba	dúmɛum bumul
‘mumble’	búrɛʔur bos	búrɛʔur war	búrɛʔur bo	búrɛʔur ba	búrɛʔurt'i
‘yell’	q'ɪɛ bos	q'awar	q'ɪɛ bo	q'ɪɛ ba	q'ɪɛt'i
‘cry’ (baby)	a'ngáʂ bos	a'ngáʂr	a'ngáʂ bo	a'ngáʂ ba	a'ngáʂt'i
‘squeak; creak’	c'ír bos	c'arár	c'ír bo	c'ír ba	c'írt'i
‘cry’	éʔmmus	éʔmmur	éʔmmu	éʔmma	éʔmt'i
‘crack (voice)’	ɛʔwɛár bos	ɛʔwɛár war	ɛʔwɛár bo	ɛʔwɛár ba	ɛʔwɛár bumul
‘sob’	s:ínk' bos	s:ánk'ár	s:ínk' bo	s:ínk' ba	s:ínk't'i
‘croak; cry’	rás bos	rás war	rás bo	rás ba	rás bumul
‘whistle’	ʂ:wít' bos	ʂ:wít'ár	ʂ:wít' bo	ʂ:wít' ba	ʂ:wítt'i
‘whistle softly’	ʂ:wít'qʔwɛit' bos	ʂ:wít'qʔwɛit' war	ʂ:wít'qʔwɛit' bo	ʂ:wít'qʔwɛit' ba	ʂ:wít'qʔwɛitt'i

**Table 7:** Ideophonic inarticulate oral expression verbs in Archi  
(data from Chumakina et al. 2007)

(27) *to-r*    ʧʰe-kul-li            eʂm war-ʂi            d-i.  
DIST-F    be\_cold-MSD-ERG    cry    say.IPF-CV            F-COP  
‘She is crying because she is cold.’

(28) *to-r*    ʔonnol            ac'i-li            q'ɪɛ bo.  
DIST-F    woman(F)(NOM)    pain-ERG    yell    say.PF  
‘This woman cried out in pain.’

(29) *un*            daki            duna:la burɛʔur            war  
2.SG.NOM    why            a.lot            grumble    say.IPF  
‘Why do you grumble so much?’

Compound verbs denoting manners of oral expression do not necessarily have a human subject; for instance, they can describe the speech itself (ex. 30):

- (30) *piriz-li*            *su-mč'iš*            *wit*            *č'at*            *łals*  
 paralysis-ERG    hold.PF.M-COND    2.SG.GEN    word(N)    stammer  
*bo-qi.*  
 say.PF-FUT

'If you are stricken by paralysis, your speech will become impaired.'

Also, some verbs in the list are occasionally used with reference to animals (ex. 31):

- (31) *s:ink'*    *bo-li*            *i-t'u*            *baħri.*  
 sob    say.PF-CV    (N)AUX-NEG    stray\_dog  
 'The stray dog did not sob.'

### 2.3. Human speech and communication

Most verbs referring to articulate speech acts are specific to human communication, as seen in Table 8. Some of them are intransitive, as in (ex. 32–34):

- (32) *zon*            *to-w-mu-łu*            *kino-li-n*            *ħaq'liq<sup>s</sup>*    *ba<sup>s</sup>bu.*  
 1SG.NOM    DIST-M-OBL-COM    film-OBL-GEN    about    talk + say.PF  
 'I was talking to him about films.'

- (33) *to-w*    *t'inna*            *č'ak'ar.*  
 DIST-M    a.little    chatter.IPF  
 'He is a bit of a chatterbox.'

- (34) *to-w*            *š:<sup>w</sup>it'q<sup>w</sup>it' + war-ši*            *qecašecer-ši.*  
 DIST-M(NOM)    whistle + say.IPF-CV    walk.IPF-CV  
 'He was walking up and down, whistling softly.'

Others are transitive, in that their construction has an agent in the ergative. The nominative patientive argument can be the coverb itself, treated as a speech act (ex. 35 and 36):



- (35) *to-w-mu za:r-ši šuš bo.*  
 DIST-M-ERG 1.OBL-CONT-ALL whisper say.PF  
 ‘He whispered to me.’

- (36) *zari š<sup>w</sup>it’ bo-qi.*  
 1.ERG whistle say.PF-FUT  
 ‘I will whistle.’

In other transitive verbs of the same semantic class, the coverb is coalescent and speech content (ex. 37) or an animate addressee (ex. 38–40) can be treated as the direct object in the nominative case:

- (37) *zari marči g<sup>w</sup>abq:’u-timme-s lagum χabu.*  
 1.ERG everybody HPL.gather.PF-ATTR.PL-DAT song(N) sing + say.PF  
 ‘I sang a song to everyone who had come.’

- (38) *nen marči zab bo oq-li-ti-k.*  
 1.PL.EXCL.ERG everybody invite say.PF wedding-OBL-SUPER-LAT  
 ‘We invited everybody to the wedding.’

- (39) *um-mu lo-bur ħat bo.*  
 father-ERG child-PL.NOM scold say.PF  
 ‘The father scolded his children.’

- (40) *zari to-w aʔ bo s:afat wiklit’u-mi-t.*  
 1.ERG DIST-M call say.PF hour(N) seven.N-OBL-SUPER  
 ‘I invited him for seven o’clock.’

- (41) *w-is halmaχ-mu zon č’elle-ši aʔ bo.*  
 M-1SG.GEN friend-ERG 1SG.NOM outside-ALL call say.PF  
 ‘My friend asked me to go out.’

The alternative construction attested in (ex. 42) is probably calqued on Russian (*zvonit’* + DAT):

- (42) *zari anži-l-aš tu-w-mi-ra-k aʔ + bo.*  
 1.ERG Makhachkala-OBL-INEL DIST-M-OBL-CONT-LAT call + say.PF  
 ‘I called him from Makhachkala.’

	PF	IPF	INF	IMPER	MSD
‘talk’	<i>báʔbus</i>	<i>báʔbur</i>	<i>báʔbu</i>	<i>báʔba</i>	<i>báʔbatʔi</i>
‘chatter’	<i>čʔákʔ bos</i>	<i>čʔakʔár</i>	<i>čʔákʔ bo</i>	<i>čʔákʔ ba</i>	<i>čʔákʔtʔi</i>
‘brag’	<i>žírq bos</i>	<i>žarqár</i>	<i>žírq bo</i>	<i>žírq ba</i>	<i>žírq bumul</i>
‘gossip’	<i>šúšqʔus bos</i>	<i>šúšqʔus war</i>	<i>šúšqʔus bo</i>	<i>šúšqʔus ba</i>	<i>šúšqʔus bumul</i>
‘whisper’	<i>šúš bos</i>	<i>šušár</i>	<i>šúš bo</i>	<i>šʔwíš ba</i>	<i>šʔwíštʔi</i>
‘sing’/‘play’	<i>χábus</i>	<i>χar</i>	<i>χábu</i>	<i>χába</i>	<i>χátʔi</i>
‘scold’	<i>ħát bos</i>	<i>ħátar</i>	<i>ħátbo</i>	<i>ħátba</i>	<i>ħátbumul</i>
‘refuse’	<i>ú:tʔu bos</i>	<i>ú:tʔu war</i>	<i>ú:tʔu bo</i>	<i>ú:tʔu ba</i>	<i>ú:tʔu bumul</i>
‘invite’	<i>záp:os</i>	<i>záb war</i>	<i>záp:o</i>	<i>záp:a</i>	<i>záp:umul</i>
‘say sharply’	<i>qʔwántʔ bos</i>	<i>qʔuntʔár</i>	<i>qʔwántʔ bo</i>	<i>qʔwántʔ ba</i>	<i>qʔwántʔ bumul</i>
‘forbid’	<i>árdigi bos</i>	<i>árdig war</i>	<i>árdig bo</i>	<i>árdigbo</i>	<i>árdig bumul</i>
‘call, invite’	<i>áʔ bos</i>	<i>aʔár</i>	<i>áʔbo</i>	<i>áʔba</i>	<i>áʔtʔi</i>

Table 8: Ideophonic speech act verbs in Archi  
 (data from Chumakina et al. 2007)

#### 2.4. Non-expressive oral and body sounds

Non-spoken sounds, given in Table 9, produced by the mouth and the nose are spontaneous and usually involuntary, and although they may be interpreted as such, they do not serve a primarily expressive function, for instance the verbs for ‘breathe’ (loudly), ‘yawn’, or ‘grit one’s teeth’. Most of these predicates also form compounds with ‘say’ (ex. 43 to 49):

- (43) *un jamu kʔellejwu haʔhar daki ?*  
 2.SG.NOM PROX.M wholly.M breathe.say.IPF why  
 ‘Why are you panting so much?’

- (44) *un daki hakʔar-ši w-i?*  
 why 2SG.NOM yawn.say.IPF-CV M-COP  
 ‘Why are you yawning?’

- (45) *to-w q'arš:ar-ši ewdi.*  
 DIST-M grit.say.IPF-CV M.be.PST  
 'He was gritting his teeth (with anger).'
- (46) *to-w iqna a'nša'r-ši ewdi.*  
 DIST-M all\_ay sneeze.say.IPF-CV M.COP.PAST  
 'He was sneezing all the time.'
- (47) *to-w nibk'i-t χ:ank'ar.*  
 DIST-M sleep.OBL-SUPER snore.say.IPF  
 'He snores in his sleep.'
- (48) *ak:ommi-tiš χaraxši is lagi q'urar-ši*  
 dawn.OBL-SUPEREL since (N)my stomach(N) rumble.say.IPF-CV  
*i.*  
 (N)COP  
 'My stomach has been rumbling since this morning.'
- (49) *b-is gon č'irq' bo.*  
 A-my finger(A) cracksay.PF  
 'My finger joint cracked.'

Some of these verbs can take a voluntary agent, via a dedicated causative construction with the auxiliary 'do', as in (ex. 50):

- (50) *to-wmu gon-nor č'irq' + bo-s ar.*  
 DIST-M.ERG finger-PL.NOM crack + say.INF (N)DO.IPF  
 'He cracks his knuckles.'

	PF	IPF	INF	IMPER	MSD
'breathe'	<i>há'h bos</i>	<i>ha'hár</i>	<i>há'h bo</i>	<i>há'h ba</i>	<i>há'ht'i</i>
'burp'	<i>řóp' bos</i>	<i>řóp' war</i>	<i>řóp' bo</i>	<i>řóp' ba</i>	<i>řóp't'i</i>
'belch; hiccough'	<i>hérq<sup>s</sup> bos</i>	<i>hérq<sup>s</sup>er</i>	<i>hérq<sup>s</sup> bo</i>	<i>hérq<sup>s</sup> ba</i>	<i>hérq<sup>s</sup>t'i</i>
'pant, puff'	<i>há'hčič' bos</i>	<i>ha'hčič'ár</i>	<i>há'hčič' bo</i>	<i>há'hčič' ba</i>	<i>há'hčič' bumul</i>
'cough'	<i>uńú bos</i>	<i>uńúr</i>	<i>uńú bo</i>	<i>uńú ba</i>	<i>uńút'i</i>
'gnash, grit (teeth)'	<i>q'árs' bos</i>	<i>q'árs'ar</i>	<i>q'árs' bo</i>	<i>q'árs' ba</i>	<i>q'árs't'i</i>
'clear one's throat'	<i>áč-bos</i>	<i>áč-bontur</i>	<i>áč-bontubu</i>	<i>áč-ballituba</i>	<i>áč-bumul</i>
'yawn'	<i>hákł' bo</i>	<i>hakł'ár</i>	<i>hákł' bos</i>	<i>hákł' ba</i>	<i>hákł' bumul</i>
'sneeze'	<i>a<sup>s</sup>nšá<sup>s</sup> bos</i>	<i>a<sup>s</sup>nšá<sup>s</sup> war</i>	<i>a<sup>s</sup>nšá<sup>s</sup> bo</i>	<i>a<sup>s</sup>nšá<sup>s</sup> ba</i>	<i>a<sup>s</sup>nšá<sup>s</sup> bumul</i>
'snore'	<i>χ:ánk' bos</i>	<i>χ:ánk'ár</i>	<i>χ:ánk' bo</i>	<i>χ:ánk' ba</i>	<i>χ:ánk' bumul</i>
'resound'	<i>dé:j bos</i>	<i>dí:jár</i>	<i>dé:j bo</i>	<i>dé:j ba</i>	<i>dé:j bumul</i>
'crack' (joints)	<i>č'írq' bos</i>	<i>č'írq'ár</i>	<i>č'írq' bo</i>	<i>č'írq' ba</i>	<i>č'írq' t'i</i>
'split, crack' (wood, skin)	<i>pírq' bos</i>	<i>pírq' war</i>	<i>pírq' bo</i>	<i>pírq' ba</i>	<i>pírq't'i</i>
'rumble' (stomach)	<i>q<sup>ws</sup>ár bos</i>	<i>q<sup>s</sup>urár</i>	<i>q<sup>ws</sup>ár bo</i>	<i>q<sup>ws</sup>ár ba</i>	<i>q<sup>ws</sup>árt'i</i>
'trow up'	<i>χáq<sup>s</sup> bos</i>	<i>χaq<sup>s</sup>ár</i>	<i>χáq<sup>s</sup> bo</i>	<i>χáq<sup>s</sup> ba</i>	<i>χáq<sup>s</sup>t'i</i>
'fart'	<i>c'íq' bos</i>	<i>c'aq'ár</i>	<i>c'íq'bo</i>	<i>c'íq' ba</i>	-
'shit'	<i>qíp:us</i>	<i>qapár</i>	<i>qíp:u</i>	<i>qíp:a</i>	<i>qípt'i</i>

**Table 9:** Ideophonic involuntary body noise verbs in Archi  
(data from Chumakina et al. 2007)

## 2.5. Voluntary body noise verbs

Voluntary noise verbs, given in Table 10, like 'spit', have their subject in the ergative case (ex. 51):

- (51) *zari č'ele-li-s tubu.*  
 1.ERG stone-OBL-DAT spit + say.PF  
 'I spat on a stone.'

The sole argument of the verb ‘blow’ can be inanimate, and then takes the nominative case, as in (ex. 52):

- (52) *hawa hubu.*  
 wind(N) blow + say.PF  
 ‘The wind blew.’

If this verb takes an object, it is an indirect one, in the dative case, but the animate subject remains marked with the ergative case (ex. 53). This non-canonical valency type results from the incorporation of the ideophone *hu*:

- (53) *zari čaraχ-li-s / oc’-li-s hubu.*  
 1.ERG lamp-OBL-DAT fire-OBL-DAT blow + say.PF  
 ‘I blew the lamp out / on the fire.’

The same valency is found with other voluntary noise verbs (ex. 54 and 55):

- (54) *zari to-w-mu-s eχ<sup>ʷ</sup>u-k ba? bo.*  
 1.ERG DIST-M.-OBL-DAT cheek.SG.IN-LAT kiss say.PF  
 ‘I kissed her on the cheek.’

- (55) *laha aq-ur-če-łu č<sup>w</sup>arχ<sup>s</sup> bo.*  
 child.ERG foot-PL-OBL.PL-COM stamp say.PF  
 ‘The child stamped his feet.’

	PF	IPF	INF	IMPER	MSD
‘blow’ (wind; person)’	<i>húbus</i>	<i>hur</i>	<i>húbu</i>	<i>húba</i>	<i>hút’i</i>
‘spit’	<i>tú bus</i>	<i>tur</i>	<i>tú bus</i>	<i>tú ba</i>	<i>tút’i</i>
‘kiss’	<i>p’á? bos</i>	<i>p’a?ár</i>	<i>p’á? bo</i>	<i>p’á? ba</i>	<i>p’á?t’i</i>
‘kiss’	<i>bá? bos</i>	<i>bá? war</i>	<i>bá? bo</i>	<i>bá?-ba</i>	<i>bá? bumul</i>
‘blow one’s nose’	<i>ú<sup>nš</sup> bos</i>	<i>u<sup>nš</sup>á<sup>r</sup></i>	<i>ú<sup>nš</sup> bo</i>	<i>ú<sup>nš</sup> ba</i>	<i>ú<sup>nš</sup>t’i</i>
‘stamp, clap’	<i>č<sup>w</sup>árχ<sup>s</sup> bos</i>	<i>čurχ<sup>s</sup>ár</i>	<i>č<sup>w</sup>árχ<sup>s</sup> bo</i>	<i>č<sup>w</sup>árχ<sup>s</sup> ba</i>	<i>č<sup>w</sup>árχ<sup>s</sup>t’i</i>

Table 10: Ideophonic voluntary body noise verbs in Archi (data from Chumakina et al. 2007)

## 2.6. Non-body-related sound verbs

Sound verbs that do not involve any human or animal referent, given in Table 11, are nevertheless frequently formed with the light verb ‘say’. They are, of course, monovalent intransitives, with their sole argument in the nominative case, as in (ex. 56–58):

(56) *b-is dump č’a’p:u.*  
 A-my ball(A) burst + say.PF  
 ‘My ball burst.’

(57) *kabk č’a’p:u-li*  
 glass(N) crack + say.PF-EVID  
 ‘The glass (window) cracked.’

(58) *dunil-li-n qu-t’i č’a’p-t’i i.*  
 sky-OBL-GEN thunder-MSD strike-MSD (N)COP  
 ‘The thunder resounded.’

	PF	IPF	INF	IMPER	MSD
‘squelch (water in boots)’	<i>wé’lč’ bos</i>	<i>wé’lč’ér</i>	<i>wé’lč’ bo</i>	<i>wé’lč’ ba</i>	<i>wé’lč’t’i</i>
‘eddy, babble, flow’ (spring)	<i>k’úrk’ur bos</i>	<i>k’úrk’ur war</i>	<i>k’úrk’ur bo</i>	<i>k’úrk’ur ba</i>	<i>k’úrk’urt’i</i>
‘drip’	<i>k’ent’ bos</i>	<i>k’ent’ér</i>	<i>k’ent’ bo</i>	<i>k’ent’ ba</i>	<i>k’éntt’i</i>
‘burst, explode; shoot’	<i>č’a’p:us</i>	<i>č’a’p’ár</i>	<i>č’a’p:u</i>	<i>č’a’p:a</i>	<i>č’a’p’t’i</i>
‘exchange fire’	<i>č’a’p’-q’ap’ bos</i>	<i>č’a’p’-q’ap’- bar</i>	<i>č’a’p’-q’ap’- bo</i>	<i>č’a’p’-q’ap’- ba</i>	<i>č’a’p’-q’ap’t’i</i>
‘sizzle’	<i>č’f’ bos</i>	<i>č’a’ár</i>	<i>č’f’ bo</i>	<i>č’f’ ba</i>	<i>č’f’ bumul</i>
‘blow (trumpet)’	<i>dú’t’ bos</i>	<i>du’t’ár</i>	<i>dú’t’ bo</i>	<i>dú’t’ ba</i>	<i>dú’t’ bumul</i>

**Table 11:** Ideophonic non-body-related noise verbs in Archi (data from Chumakina et al. 2007)

### 3. Non-sound verbs

The use of the verb ‘say’ in Archi verb compounds extends far beyond the domain of speech and noises. This section deals with predicates denoting events and actions that are not primarily sounds, nor are they produced by the vocal tract.

#### 3.1. Ingestion verbs

Ideophonic compounds cover all the range of ingestion verbs, given in Table 12, including two verbs for ‘drink’ (cold vs hot drinks), with the exception of the two Archi verbs for ‘eat’ (for which see section 9). Digestion verbs are transitive (ex. 59–64):

(59) *zari semečka čʷečʷə bo.*

1.ERG seed(A) nibble say.PF

‘I nibbled on sunflower seeds.’

(60) *laha gon-nor lamar-ši i.*

child.ERG finger-PL.NOM lick + say.IPF-CV (M)COP

‘The child licks its fingers.’

(61) *b-olo gʷači-li leki bʷemʷer.*

HPL-1.PL.GEN dog-ERG bone(N) gnaw + say.IPF

‘Our dog is gnawing at the bone.’

(62) *zari ak:ommiš čij χ:ur.*

1.ERG dawn.DAT tea(N) drink + say.IPF

‘In the morning I drink tea.’

(63) *zari lap izu-t diqʷ χ:ubu.*

1.ERG very tasty.ATTR-N soup(N) drink + say.PF

‘I had some very nice soup.’

- (64) χ<sup>ʰ</sup>elemi-s            harak      diq’            χurk’ar-gi!  
 guest.OBL.PL-DAT    in\_front    soup(N)    slurp.say.IPF-PROHIB  
 ‘Do not slurp your soup when guests are around.’

	PF	IPF	INF	IMPER	MSD
‘gorge’	šá <sup>ʰ</sup> mmus	ša <sup>ʰ</sup> má <sup>ʰ</sup> r	šá <sup>ʰ</sup> mmu	šá <sup>ʰ</sup> mma	šá <sup>ʰ</sup> mšmul
‘eat food made of dac’on’	χ <sup>ʰ</sup> áχ <sup>ʰ</sup> :u bos	χ <sup>ʰ</sup> áχ <sup>ʰ</sup> :or	χ <sup>ʰ</sup> áχ <sup>ʰ</sup> :u bo	χ <sup>ʰ</sup> áχ <sup>ʰ</sup> :u ba	χ <sup>ʰ</sup> áχ <sup>ʰ</sup> :u bumul
‘gnaw’ (dogs)	ʋ <sup>ʰ</sup> émʋ <sup>ʰ</sup> ə bos	ʋ <sup>ʰ</sup> émʋ <sup>ʰ</sup> er	ʋ <sup>ʰ</sup> émʋ <sup>ʰ</sup> ə bo	ʋ <sup>ʰ</sup> émʋ <sup>ʰ</sup> ə ba	ʋ <sup>ʰ</sup> émʋ <sup>ʰ</sup> ət’i
‘overeat (animals), glut (people)’	t’ént’ə bos	t’ént’er	t’ént’ə bo	t’ént’ə ba	t’ént’ə bumul
‘nibble’	č <sup>ʷ</sup> éč’ə bos	č <sup>ʷ</sup> eč’ér	č <sup>ʷ</sup> éč’ə bo	č <sup>ʷ</sup> éč’ə ba	č <sup>ʷ</sup> éč’ə bumul
‘chew’	č’á <sup>ʰ</sup> mmus	č’á <sup>ʰ</sup> má <sup>ʰ</sup> r	č’á <sup>ʰ</sup> mmu	č’á <sup>ʰ</sup> mma	č’á <sup>ʰ</sup> mt’i
‘champ, chew with open mouth’	č’ú <sup>ʰ</sup> w bos	č’ú <sup>ʰ</sup> w war	č’ú <sup>ʰ</sup> w bo	č’ú <sup>ʰ</sup> w ba	č’ú <sup>ʰ</sup> wt’i
‘eat too little, unwillingly’	sém bos	sém war	sém bo	sém ba	sémt’i
‘swallow without chewing’	gúr <sup>q</sup> ’s bos	gur <sup>q</sup> ’ár	gúr <sup>q</sup> ’s bo	gúr <sup>q</sup> ’s ba	gúr <sup>q</sup> ’t’i
‘graze’	há <sup>ʰ</sup> nč’ bos	ha <sup>ʰ</sup> nč’ár	há <sup>ʰ</sup> nč’ bo	há <sup>ʰ</sup> nč’ ba	há <sup>ʰ</sup> nč’t’i
‘lick’	lammus	lamár	lámmu	lámma	lámmul
‘lick clean’ (plates)	čák’ bos	čak’ár	čák’ bo	čák’ ba	čák’t’i
‘slurp’	χ <sup>ʰ</sup> úr <sup>k</sup> ’ bos	χ <sup>ʰ</sup> ur <sup>k</sup> ’ár	χ <sup>ʰ</sup> úr <sup>k</sup> ’ bo	χ <sup>ʰ</sup> úr <sup>k</sup> ’ ba	χ <sup>ʰ</sup> úr <sup>k</sup> ’t’i
‘sip smth hot’	χ <sup>ʰ</sup> úbus	χ <sup>ʰ</sup> ur	χ <sup>ʰ</sup> úbu	χ <sup>ʰ</sup> úba	χ <sup>ʰ</sup> út’i
‘drink (cold)’	c’ábus	c’ár	c’ábu	c’á ba	c’át’i

Table 12: Ideophonic ingestion verbs in Archi (data from Chumakina et al. 2007)



### 3.2. Sensation verbs

Archi verbs denoting non-auditive sensations are also ideophonic compounds. The compound verbs for ‘smell’, given in Table 13, are ambitransitive (ex. 65):

(65) *a. s:unt’ bo-li ha<sup>h</sup>mp bo-li.*  
 smell say.PF-CV bark say.PF-CV  
 ‘sniffing and barking’

*b. zari t’e<sup>s</sup>-tu s:unt’ bo.*  
 1.ERG flower-PL.NOM smell say.PF  
 ‘I smelled the flowers.’

Meanwhile all other verbs of this semantic class are intransitive (ex. 66–74):

(66) *dunil wak’ bo-li dati.*  
 sky(N) clear say.PF-CV clear.PF  
 ‘The sky cleared.’

(67) *dunil par bo.*  
 sky(N) flash say.PF  
 ‘Lightning flashed (lit. ‘the sky flashed’).’

(68) *oc’ pirχ bo.*  
 fire(N) blaze say.PF  
 ‘The fire blazed up.’

(69) *χ:<sup>w</sup>alli pirχ bo.*  
 bread(A) blaze say.PF  
 ‘The bread is slightly burnt.’

(70) *is aq guk:or-ši i.*  
 (N)my foot(N) itch.say.IPF-CV (N)COP  
 ‘My foot is itching.’

(71) *is aq c'urar-ši i.*  
 (N)my foot(N) ache.say.IPF-CV COP.N  
 'My leg aches.'

(72) *s:ol lah bo-li imc:'-i-s:-u*  
 fox(NOM) hunger say.PF-EVID honey-OBL-DAT-AND  
*inχ:-i-s:-u χir.*  
 butter-OBL-DAT-AND behind  
 'The fox was craving for (lit. behind) more honey and butter.'

(73) *zon ja-t ari a-s ħuč' bo.*  
 1SG.NOM PROX-N work(N) (N)do-INF be.lazy say.PF  
 'I was too lazy to do this job.'

(74) *jas:ana uχ k'ij bo-li.*  
 this.year field(N) overripe say.PF-CV  
 'This year our wheat was overripe...'

	PF	IPF	INF	IMPER	MSD
'smell'	<i>súnt' bos</i>	<i>sunt'ar</i>	<i>súnt' bo</i>	<i>súnt' ba</i>	<i>súntt'i</i>
'smoke; smell'	<i>p'áh bos</i>	<i>p'ahár</i>	<i>p'áh bo</i>	<i>p'áh ba</i>	<i>p'áht'i</i>
'itch'	<i>χ<sup>ws</sup>ár bos</i>	<i>χ<sup>i</sup>urár</i>	<i>χ<sup>ws</sup>ár bo</i>	<i>χ<sup>ws</sup>ár ba</i>	<i>χ<sup>ws</sup>árt'i</i>
'be overdone' (meat)	<i>túr bos</i>	<i>turár</i>	<i>túr bo</i>	<i>túr ba</i>	<i>túr bumul</i>
'become sour'	<i>q'ánc' bos</i>	<i>q'anc'ár</i>	<i>q'ánc' bo</i>	<i>q'ánc' ba</i>	<i>q'ánc't'i</i>
'ache'	<i>c'úk' bos</i>	<i>c'úk' war</i>	<i>c'úk' bo</i>	<i>c'úk' ba</i>	<i>c'úk' bumul</i>
'throb (pain)'	<i>wá'rt bos</i>	<i>wa'rtár</i>	<i>wá'rt bo</i>	<i>wá'rt ba</i>	<i>wá'rt bumul</i>
'want to eat smth'	<i>láhbos</i>	<i>lahár</i>	<i>láhbo</i>	<i>láhba</i>	<i>láht'i</i>
'be lazy, not hungry'	<i>ħúč' bos</i>	<i>ħúč' war</i>	<i>ħúč' bo</i>	<i>ħúč' ba</i>	<i>ħúč't'i</i>

Table 13: Sample of deophonic sensation verbs in Archi  
 (data from Chumakina et al. 2007)

### 3.3. Intransitive movement verbs

Many intransitive movement verbs, as those given in Table 14, expressed by ideophonic compounds have inanimate subjects, as in (ex. 75–83):

- (75) *bejraq parar-ši i.*  
 flag(N) fly + say.IPF-CV (N)COP  
 ‘The flag is flying.’
- (76) *ha<sup>h</sup>təra čax<sup>h</sup>:ar-ši i.*  
 river(N) flow.say.IPF-CV (N)COP  
 ‘The river flows.’
- (77) *dump gugo:bo.*  
 ball(a) roll + say.pf  
 ‘The ball rolled (away).’
- (78) *harq čax<sup>h</sup>:ar.*  
 roof(N) leak.say.IPF  
 ‘The roof leaks.’
- (79) *deq<sup>h</sup> čirχ<sup>h</sup> bo-li.*  
 road(A) collapse say.PF-CV  
 ‘The road was washed away...’
- (80) *mišin-ni-t da<sup>h</sup>nka<sup>h</sup>r-ši i.*  
 car-OBL-SUPER shake.say.IPF-CV (N)COP  
 ‘The car is shaking.’
- (81) *nok<sup>h</sup> dimmu.*  
 house(N)(NOM) collapse + say.PF  
 ‘The house got destroyed.’
- (82) *sot ga<sup>h</sup>rtar-ši i.*  
 tooth(N) be.loose.say.IPF-CV (N)COP  
 ‘The tooth is loose.’

- (83) *to-w-mu-n*      *aq*      *q<sup>s</sup>art'*      *bo-li*  
 DIST-M-OBL-GEN    foot(N)    dislocate    say.PF-CV  
 'He dislocated his foot.'

However, animal (ex. 84 and 85) or human subjects (ex. 86–94) are also well represented:

- (84) *no<sup>s</sup>š*      *pars*      *bo.*  
 horse(A)    start    say.PF  
 'The horse gave a start (in fright).'

- (85) *parχ*    *bo-li*      *oq<sup>a</sup>-li*      *noc'*      *laha-rak.*  
 fly    say.PF-CV    (N)leave.PF-CV    bird(N)    child.OBL-CONTLAT  
 'The bird flew to the child.'

- (86) *nen*      *dimmu-li*      *nołdorčej-ši*      *oq<sup>a</sup>.*  
 1.PL.NOM.EXCL    disperse.say.PF-CV    house.PL.IN-ALL    leave.PF.1.PL  
 'We went off to our homes.'

- (87) *olo*      *lo*      *sanna*      *łurχ<sup>s</sup>ar-ši-t'aw*      *i-t'u.*  
 (N)our    child(N)    so.far    crawl.say.IPF-CV-NEG.CV    (N)COP-NEG  
 'Our child is still only crawling (lit. is not yet not crawling).'

- (88) *zon*      *akrommis*      *duwraz-l-a*      *t'inna*      *qecšec*      *bo.*  
 1SG.NOM    dawn.DAT    gate-OBL-IN    a.little    walk    say.PF  
 'I walked a bit in the yard this morning.'

- (89) *zon*      *qer-ši*      *w-i.*  
 1SG.NOM    dance.say.IPF-CV    M-COP  
 'I am dancing.'

- (90) *duχ:ul-li-t:i-k*      *askar*      *t'ank' + bo-li.*  
 village-OBL-SUPER-LAT    squad    jumpsay.PF-EVID  
 'A squad had attacked the village.'

- (91) *un os t'inna t'irk' ba!*  
 2SG.nom one a.little speed.up say.IMPER  
 'Speed up a bit, we will be late.'
- (92) *mu-ši q'owq'i žular-gi!*  
 fine-ADV sit.M.IMPER fidget.say.IPF-PROHIB  
 'Sit still, don't fidget!'
- (93) *imi-ši eši dur bo-li q<sup>w</sup>ali*  
 there-ALL here walk.fast say.PF-CV (M)COME.PF-EVID  
 'He came here from over there walking fast.'
- (94) *dija q'or-če-łtu kušar-ši w-i.*  
 father(NOM) book.PL-OBL.PL-COM rummage.IPF-CV M-COP  
 'Father is rummaging through books.'

	PF	IPF	INF	IMPER	MSD
'fidget, swarm'	ž <sup>w</sup> ál bos	žulár	ž <sup>w</sup> ál bo	ž <sup>w</sup> ál ba	ž <sup>w</sup> ált'i
'crash, plump'	wá'p' bos	wá'p' war	wá'p' bo	wá'p' ba	wá'p bumul
'be loose' (tooth)	gá <sup>r</sup> rtbos	ga <sup>r</sup> rtá <sup>r</sup>	gá <sup>r</sup> rtbo	gá <sup>r</sup> rtba	gá <sup>r</sup> rtbumul
'roll out'	gér <sup>g</sup> i bos	gér <sup>g</sup> er	gér <sup>g</sup> ibo	gér <sup>g</sup> i ba	gér <sup>g</sup> i bumul
'rock' (cot)	gér <sup>k</sup> :ibos	gér <sup>k</sup> :er	gér <sup>k</sup> :ibo	gér <sup>k</sup> :i ba	gér <sup>k</sup> :ə bumul
'go curly' (hair)	k'é <sup>r</sup> k'ə bos	k'é <sup>r</sup> k'er	k'é <sup>r</sup> k'ə bo	k'é <sup>r</sup> k'ə ba	k'é <sup>r</sup> k'ət'i
'come apart' (hay; road)	čír <sup>χ</sup> bos	čír <sup>χ</sup> war	čír <sup>χ</sup> bo	čír <sup>χ</sup> ba	čír <sup>χ</sup> bumul
'get stuck in marshy soil'	pú <sup>r</sup> š bos	pu <sup>r</sup> šá <sup>r</sup>	pú <sup>r</sup> š bo	pú <sup>r</sup> š ba	pú <sup>r</sup> š bumul
'dance'	qébus	qer	qébu	qé ba	qét'i
'jump'	t'ánk' bos	t'ánk'ár	t'ánk' bo	t'ánk' ba	t'ánk't'i
'crawl'	łúr <sup>χ</sup> bos	łur <sup>χ</sup> ár	łúr <sup>χ</sup> bo	łúr <sup>χ</sup> ba	łúr <sup>χ</sup> t'i
'shuffle one's feet'	łér <sup>χ</sup> bos	łur <sup>χ</sup> ár	łér <sup>χ</sup> bo	łér <sup>χ</sup> ba	łur <sup>χ</sup> t'i
'fly'	pár <sup>χ</sup> bos	par <sup>χ</sup> ár	pár <sup>χ</sup> bo	pár <sup>χ</sup> ba	pár <sup>χ</sup> t'i

**Table 14.** Sample of ideophonic intransitive movement verbs in Archi  
 (data from Chumakina et al. 2007)

### 3.4. Posture and attitude verbs

A small class of intransitives denote a change of state or a stative predicate (cf. Table 15, and ex. 95–100):

- (95) *to-w*      *q'ač'-bo-tu-b*                      *kuc*      *b-aku-ra?*  
 DIST-M      dress.up-say.PF-ATTR-A      way(A)      A-see.PF-Q  
 'Can you see the way he dressed up?'

- (96) *q<sup>w</sup>at'i-li-tiš*      *kł'ara:ši*      *ja<sup>t</sup>'i*      *gerk:er-ši*                      *b-i.*  
 tree-OBL-SUPEREL      down.there      snake(A)      hang.over + SAY.IPF-CV      A-COP  
 'A snake is hanging down from the tree.'

- (97) *zon*      *t'inna*      *k'ur + bo-li*      *erdi.*  
 1SG.NOM      a.little      lay + say.PF-CV      F.BE.PST  
 'I lay down for a short while.'

- (98) *ja-t*      *lo*      *k'ur + war-ši*      *i-t'u.*  
 PROX-N      child(N)      sleep + say.IPF-CV      (N)COP-NEG  
 'This child does not go to sleep easily.'

- (99) *wa-ł:ut:-ib*                      *halmaχ-til*      *q'ač + bo-na*                      *b-erq<sup>ε</sup>-u-qi.*  
 2.OBL-COM-ATTR.PL      friend-PL      array + say.PF-CV      HPL-walk-PF-FUT  
 'Your friends will go dressed up.'

- (100) *to-w*      *muq<sup>ε</sup>u*      *lobur-če-łu*                      *q'ec'er.*  
 DIST-M      small      child.PL-OBL.PL-COM      nag + say.IPF  
 'He nags little children.'

	PF	IPF	INF	IMPER	MSD
'fall asleep'	<i>k'úr k'ur bos</i>	<i>k'úr k'ur war</i>	<i>k'úr k'ur bo</i>	<i>k'úr k'ur ba</i>	<i>k'úr k'ur t'i</i>
'take a nap'	<i>k'úr bos</i>	<i>k'urár</i>	<i>k'úr bo</i>	<i>k'úr ba</i>	<i>k'úrt'i</i>
'measure one's strength with smb'	<i>qat' bos</i>	<i>qat' war</i>	<i>qat' bo</i>	<i>qat' ba</i>	<i>qat' bumul</i>
'dress up'	<i>q'áč' bos</i>	<i>q'ač'ár</i>	<i>q'áč' bo</i>	<i>q'áč' ba</i>	<i>q'áč't'i</i>
'have doubts'	<i>wás bos</i>	<i>wasár</i>	<i>wás bo</i>	<i>wás ba</i>	<i>wás bumul</i>
'nag, bother'	<i>q'éc' bos</i>	<i>q'éc'er</i>	<i>q'éc' bo</i>	<i>q'éc' ba</i>	<i>q'éc't'i</i>

**Table 15.** Ideophonic posture and attitude verbs in Archi  
(data from Chumakina et al. 2007)

### 3.5. Transitive movement verbs

Some compound verbs of motion are ambitransitive and can also have an agent in the ergative (ex. 101 and 102) (Table 16):

(101) a. *budraš na'kɬ' čʷarɤʃ bo.*  
 bucket.INEL milk(N) pour say.PF  
 'The milk poured out of the bucket.'

b. *zari ɬan čʷarɤʃ bo.*  
 1.ERG water(N) pour say.PF  
 'I poured the water out.'

(102) a. *χeme-r-ši bo χara:ši daɤ bo-li.*  
 woman.OBL.PL-CONT-ALL say.PF after turn say.PF-CV  
 'She said to the women, having turned back...'

b. *zari ɤaterte-q'a-k buɤ:i daɤ bo-qi.*  
 1SG.ERG cattle-OBL-INTER-LAT cow.PL.NOM drive say.PF-FUT  
 'I will drive the cows to the herd.'

c. *q'amatu eši tenši daɤ bo-li.*  
 hair(N) here there tousle say.PF-CV  
 '(She) tousled (her) hair.'

Other verbs are strictly transitive (ex. 103–106), although the agent can be left out (ex. 107–109) or be in the nominative case in (ex. 110), which is a case of ‘binominative construction’ denoting a continuous activity<sup>3</sup>:

- (103) *zari dump χ<sup>s</sup>ološi lapu.*  
 1.ERG ball(A) upwards throw.say.PF  
 ‘I threw the ball upwards.’
- (104) *zari q<sup>r</sup>art<sup>i</sup>-li χ<sup>w</sup>alli gergi bo.*  
 1.ERG pin-ERG bread(A) roll.out say.PF  
 ‘I rolled out bread with a rolling pin.’
- (105) *zari to-w ba<sup>s</sup>re-q<sup>a</sup>-k χit’ bo.*  
 1.ERG DIST-M lake.OBL-INTER-LAT push say.PF  
 ‘I pushed him into the lake.’
- (106) *doba-mu tung-li-tu oł tarχ:ar-ši i*  
 granny-ERG spindle-OBL-COM wool(N) twist.IPF-CV COP(N)  
*imkł’el-lur-če-s.*  
 sock-PL-PL.OBL-DAT  
 ‘The grandmother twists a woollen thread with a spindle to knit socks.’
- (107) *č’ut zerzer-gi!*  
 jug(N) fill.say.IPF-PROHIB  
 ‘Do not fill the jug to the brim.’
- (108) *olo duχriq<sup>a</sup>-š os lo χapu-li.*  
 (N)our village.SG.INTER-EL one child(F) kidnap.say.PF-CV  
 ‘Once a girl was kidnapped from our village.’
- (109) *č’abu dere bo-mχur*  
 sheep.PL separate.lambs say.PF-TEMP2  
 ‘when the ewes were separated from the lambs’

<sup>3</sup> For binominative, aka biabsolutive, constructions, cf. Chumakina and Bond (2016).



- (110) *to-r misgin eqiša kelkan č'eᵑp*  
 DIST-F poor at.night until cot(A)  
*gerkir-ši erdi.*  
 rock.say.IPF-CV F.BE.PST  
 'She was rocking the cot until the evening, poor thing.'

	PF	IPF	INF	IMPER	MSD
'turn, squeeze'	<i>q<sup>w</sup>éraq'ə bos</i>	<i>q<sup>w</sup>éraq'er</i>	<i>q<sup>w</sup>éraq'əbo</i>	<i>q<sup>w</sup>éraq'ə ba</i>	<i>q<sup>w</sup>éraq'ət'i</i>
'jam, crush, squash'	<i>ɁúrkɁ' bos</i>	<i>ɁurkɁ'ár</i>	<i>ɁúrkɁ' bo</i>	<i>ɁúrkɁ' ba</i>	<i>ɁúrkɁ't'i</i>
'push, cram, stuff in'	<i>š<sup>w</sup>élš:i bos</i>	<i>š<sup>w</sup>élš:ér</i>	<i>š<sup>w</sup>élš:i bo</i>	<i>š<sup>w</sup>élš:i ba</i>	<i>š<sup>w</sup>élš:i bumul</i>
'dig, take off a layer'	<i>qért' bos</i>	<i>qért'er</i>	<i>qért' bo</i>	<i>qért' ba</i>	<i>qértt'i</i>

**Table 16.** Sample of ideophonic transitive movement verbs in Archi  
 (data from Chumakina et al. 2007)

### 3.6. Other transitive action verbs

Among other transitive action verbs involving the light verb 'say', most can be conceived of as associated with a sound (cf. Table 17, ex. 111) or a movement (ex. 112 and 113):

- (111) *ert:i-ttur*                      *laha*      *raɣ*      *a<sup>r</sup>qu<sup>r</sup>/.*  
 F.become.PF-ATTR.F    child.ERG    clay      break.say.IPF  
*evt:i-ttu*                      *laha*      *daʎ' q<sup>w</sup>ank'*      *war.*  
 M.become.PF-ATTR    child.ERG    door slam      say.IPF  
 'A grown daughter breaks the dishes, a grown son slams the door.'

- (112) *χ<sup>w</sup>ak:a*    *nen*                      *χ:uχ:ət'iki-li*      *q<sup>w</sup>at'i*      *χ:uχ:u bo.*  
 forest.IN    1.PL.EXCL.NOM    saw-ERGt      ree(A)      saw      say.PF  
 'We cut down a tree in a forest with a saw.'

- (113) *dija-mu k'as:a is kart'i χ:art' bo.*  
 father-ERG knife.ERG (N)my head(N) shave say.PF  
 'Father shaved my head with his knife.'

	PF	IPF	INF	IMPER	MSD
'saw'	<i>χ:úχ:u bos</i>	<i>χ:úχ:or</i>	<i>χ:úχ:u bo</i>	<i>χ:úχ:u ba</i>	<i>χ:úχ:u bumul</i>
'shave'	<i>χ:árt' bos</i>	<i>χ:art'ár</i>	<i>χ:árt' bo</i>	<i>χ:árt' ba</i>	<i>χ:árt' bumul</i>
'sew with short stitches'	<i>qúc bos</i>	<i>qúc war</i>	<i>qúc bo</i>	<i>qúc ba</i>	<i>qúc bumul</i>
'twist wool'	<i>tárχ bos</i>	<i>tarχ:ár</i>	<i>tárχbo</i>	<i>tárχba</i>	<i>tárχt'i</i>
'spin' (thread, wool)	<i>qénz bos</i>	<i>qénzer</i>	<i>qénz bo</i>	<i>qénz ba</i>	<i>qénzt'i</i>
'comb threads while weaving'	<i>χ<sup>w</sup>it' bós</i>	<i>χ<sup>w</sup>it' war</i>	<i>χ<sup>w</sup>it' bó</i>	<i>χ<sup>w</sup>it' bá</i>	<i>χ<sup>w</sup>itt'i</i>
'shear sheep, trim hair'	<i>q'énc' bos</i>	<i>q'enc'ér</i>	<i>q'énc' bo</i>	<i>q'énc' ba</i>	<i>q'énc't'i</i>
'cut' (grass)	<i>β<sup>s</sup>áz bos</i>	<i>β<sup>s</sup>azár</i>	<i>β<sup>s</sup>áz bo</i>	<i>β<sup>s</sup>áz ba</i>	<i>β<sup>s</sup>ázt'i</i>
'break (crockery)'	<i>βérβə bos</i>	<i>βerβér</i>	<i>βérβə bo</i>	<i>βérβə ba</i>	<i>βérβət'i</i>
'hew, grind smooth'	<i>pálc' bos</i>	<i>palc'ár</i>	<i>pálc' bo</i>	<i>pálc' ba</i>	<i>pálc't'i</i>
'roast grain'	<i>sése bos</i>	<i>séser</i>	<i>sése bo</i>	<i>sése ba</i>	<i>séset'i</i>
'rub hard (clothes)'	<i>β<sup>s</sup>úž bos</i>	<i>β<sup>s</sup>úž war</i>	<i>β<sup>s</sup>úž bo</i>	<i>β<sup>s</sup>úž ba</i>	<i>β<sup>s</sup>úž bumul</i>
'knock'	<i>dárq bos</i>	<i>darqár</i>	<i>dárq bo</i>	<i>dárq ba</i>	<i>darqt'i</i>
'pinch'	<i>q<sup>s</sup>ént' bos</i>	<i>q<sup>s</sup>ént'er</i>	<i>q<sup>s</sup>ént' bo</i>	<i>q<sup>s</sup>ént' ba</i>	<i>q<sup>s</sup>ént' bumul</i>
'stew (nettle); sputter'	<i>pút bos</i>	<i>putár</i>	<i>pút bo</i>	<i>pút ba</i>	<i>pútt'i</i>
'plough'	<i>húj bos</i>	<i>hijár/hejár</i>	<i>húj bo</i>	<i>húj ba</i>	<i>hújt'i</i>
'wash (oneself or clothes)'	<i>čúčə bos</i>	<i>čúčor</i>	<i>čúčə bo</i>	<i>čúčə ba</i>	<i>čúčət'i</i>

**Table 17.** Ideophonic transitive action verbs in Archi  
 (data from Chumakina et al. 2007)

However, the origin of the coverbs contained in most of these compounds is not clearly ideophonic, and some may indeed have a different origin, like a coverbal form of a

basic intransitive verb used transitively<sup>4</sup> or in causative periphrasis. This is certainly the case for *gúkə bos* ‘scratch, to itch’, whose cognate in Kryz is the simplex verb *gugayc* ‘itch’, forming the causative compound *gugara aric* ‘scratch’, and of *š:ʷélš:i bos* ‘stuff, cram’, a reduplicated form which in related languages has cognate simplex verbs based on a root *š:ʷ-*.

Still some other verbs, like *pút bos* ‘stew (nettle) < ‘sputter’ or *s:ús:e bos* ‘smear’, seem clearly ideophonic, and at least the verb *hij bos* ‘plough’ appears to derive from a construction denoting a speech act. Indeed, besides the construction where the object of this verb is ‘the field’ (ex. 114a and 114b), the texts feature cases where it seems to be constructed with draught animals as the object (ex. 115; these examples are taken from the Archi text corpus, Arkhipov et al. 2007):

- (114) a. *nen*                      *ʔanna-t*                      *uχ*                      *hejar.*  
 1.PL.NOM.EXCL    spring.OBL-SUPER    field(N)    plough.say.IPF  
 ‘We plough our field in the spring.’

- b. *uχ*            *hij*            *ba!*  
 field    plough    say.IMPER  
 ‘Plough the field!’

- (115) *imi-š*            *eš:i*            *hij*            *bo-li*            *uns-ur...*  
 there-EL            here.to            “hey !”            say.PF-CV    bull-PL  
 ‘From there, having driven / \*ploughed the oxen here...’

The explanation for this unusual behaviour is provided by external comparison, as a related periphrastic construction is found in Kryz, where the verb ‘do’ is used instead of ‘say’ as a light verb. In Kryz, *haj aridž* means ‘send, let go’. In both languages, the compound contains the exclamation *haj*, used primarily to hurry up large cattle and horses. However, while in Kryz the derived meaning extends by including the horse driver as the object, in Archi the animate object (the oxen) can be left out and the resulting intransitive verb ‘drive’ is understood as ‘plough’ and can therefore take a new, inanimate object (the field).

<sup>4</sup> The important and difficult question of labile verbs, a.k.a. ambitransitives, remains to be studied in Archi as well as in most languages of its family.

#### 4. The phonetic structure of ideophonic compounds

Ideophonic compounds show some interesting morphonological processes and phonotactic features.

##### 4.1. Morphology of ‘say’ as a light verb

The light verb part of ideophonic compounds can be affected in various ways. Seven configurations can be distinguished based on the distribution of surface forms of the light verb:

- the verb ‘say’ remains intact in all principal parts of the compound;
- the verb ‘say’ remains intact in all principal parts of the compound except for the masdar, which takes the suffix borrowed from Lak (-*t’i*, instead of *bumul*);
- the verb ‘say’ remains intact in all principal parts except the imperfective;
- the verb ‘say’ surfaces as a suffix *-bus* in the infinitive;
- the verb ‘say’ surfaces as a suffix *-bus* in the infinitive, and the imperfective is modified;
- the verb ‘say’ surfaces as a suffix *-bus* in the infinitive, the imperfective is modified and the masdar is in *-t’i*;
- the verb ‘say’ surfaces as a suffix *-mus* in the infinitive.

These types are illustrated in Table 18, arranged from phonetically most stable to most divergent:

	PF	IPF	INF	IMPER	MSD
‘buzz’	<i>zíz bos</i>	<i>zíz war</i>	<i>zíz bo</i>	<i>zíz ba</i>	<i>zíz bumul</i>
‘grumble, mumble’	<i>búrɓʷur bos</i>	<i>búrɓʷur war</i>	<i>búrɓʷur bo</i>	<i>búrɓʷur ba</i>	<b><i>búrɓʷurt’i</i></b>
‘drizzle’	<i>c’éc’ bos</i>	<i>c’éc’er</i>	<i>c’éc’ bo</i>	<i>c’éc’ ba</i>	<i>c’éc’ə bumul</i>
‘bray’	<i>hóʔ bos</i>	<i>haʔár</i>	<i>hóʔ bo</i>	<i>hóʔ ba</i>	<b><i>hóʔt’i</i></b>
‘growl’	<i>ħúbus</i>	<i>ħur</i>	<i>ħúbu</i>	<i>ħú ba</i>	<b><i>ħút’i</i></b>
‘gorge’	<i>šáʷmmus</i>	<i>šaʷmáʷr</i>	<i>šáʷmmu</i>	<i>šáʷmma</i>	<i>šáʷmšmul</i>
‘cry’	<i>éʷmmus</i>	<i>éʷmmur</i>	<i>éʷmmu</i>	<i>éʷmma</i>	<b><i>éʷmt’i</i></b>

**Table 18.** Phonetic shift gradient in Archi ideophonic compounds  
(data from Chumakina et al. 2007)

These phonetic changes are observed in all semantic subcategories of ideophonic compounds, with no apparent functional correlations. For the morphemic variant *t'i*, which is a Lak borrowing, no clear motivation was found.

#### 4.2. Phonotactics of ideophonic coverbs

The phonetic structure of ideophonic coverbs is monosyllabic and conforms to any possible syllable template of Archi nouns, i.e. involving no initial clusters and no occlusive clusters. By far the most frequent template is CVC, as illustrated in Table 19:

<i>báʔ bos</i>	'kiss'	<i>pár bos</i>	'stream, fly'
<i>čř bos</i>	'sizzle'	<i>q'éc' bos</i>	'nag, bother'
<i>hóʔ bos</i>	'bray'	<i>qíp:us</i>	'defecate'
<i>k'ij bos</i>	'overripen (wheat)'	<i>qúc bos</i>	'sew with short stitches'
<i>k'úr bos</i>	'take a nap'	<i>χ<sup>w</sup>ár bos</i>	'itch'
<i>kúr bos</i>	'rub; rub oneself'	<i>χ<sup>w</sup>it' bós</i>	'comb threads while weaving'

Table 19. Some CVC Archi ideophonic coverbs  
(data from Chumakina et al. 2007)

The second most frequent template is CVRC (Table 20):

<i>bú'rt' bos</i>	'rock' (baby); 'eat one's fill'	<i>łérχ<sup>s</sup> bos</i>	'shuffle one's feet (people)'
<i>č'é'nt' bos</i>	'squeeze, milk slowly; spit through one's teeth'	<i>ní'w bos</i>	'meow'
<i>č<sup>w</sup>árχ<sup>s</sup> bos</i>	'stamp, clap'	<i>q<sup>w</sup>ank' bos</i>	'slam' (door)
<i>dá'nk bos</i>	'pulse' (heart)	<i>žír q bos</i>	'brag'
<i>dá'nk bos</i>	'be shaken, shake'	<i>gúrq<sup>s</sup> bos</i>	'swallow without chewing'
<i>dár q bos</i>	'knock'	<i>há'nč' bos</i>	'graze'
<i>gá'rt bos</i>	'be loose' (tooth)	<i>χ:úrk' bos</i>	'slurp'
<i>ha'mp bos</i>	'bark'	<i>súnt' bos</i>	'smell'
<i>łáls bos</i>	'stammer'	<i>zímł bos</i>	'break out in a rash'
		<i>zámq' bos</i>	'become very cold (weather)'

Table 20. Some CVRC Archi ideophonic coverbs  
(data from Chumakina et al. 2007)

Other existing templates are shown in Table 21:

CV	<i>q'ėbus</i>	'butt; stick in, thrust'		<i>uħú bos</i>	'cough'
	<i>c'ábus</i>	'drink (cold)'	CVCV	<i>hačá bos</i>	'carry on one's back'
VC	<i>áχ bos</i>	'clear throat and spit'		<i>deré bos</i>	'separate lambs'
	<i>úw bos</i>	'howl'	VRC	<i>ú'nš bos</i>	'blow one's nose'
VCV	<i>úrt'u bos</i>	'refuse'	CVCVR	<i>dítan bos</i>	'caress' (a child)

**Table 21.** Some Archi ideophonic coverbs with untypical templates  
(data from Chumakina et al. 2007)

The remaining templates are cases of reduplication or borrowed words.

### 4.3. Reduplication

A restricted number of Archi ideophones involve reduplication and exhibit a range of different shapes, as can be seen in Table 22:

CVC	<i>c'éc' bos</i>	'drizzle'
	<i>víx bos</i>	'moo, bellow'
CVCV	<i>ħiħi bos</i>	'whinny'
	<i>sú:se bos</i>	'coat, smear'
CVRCV	<i>t'ént'ə bos</i>	'overeat (animals), glut (people)'
	<i>gérgi bos</i>	'roll out'
CVRCVR	<i>k'úrk'ur bos</i>	'purl, babble, flow' (spring); 'fall asleep'
	<i>dúmæum bos</i>	'mumble, grumble'
others	<i>šp̣q'ip bos</i>	'hurry sheep up'
	<i>urí'ʔiʔu' bos</i>	'crow (rooster)'

**Table 22.** Some Archi ideophonic coverbs with reduplication  
(data from Chumakina et al. 2007)

## 5. Derivaton from ideophonic verbs

Ideophones acquire, via the light verb 'say', the ability to form lexicalized participles, as well as a few causatives or pseudo-causatives.

### 5.1. Ideophonic adjectives from participles

In principle, any ideophonic verb can form (perfective) participles by adding attributive endings to the stem. For instance, the concrete, originally ideophonic meaning of *áʔ bos* ‘invite’ is preserved in its derivate *áʔbotut*, which means not only ‘invited’ but also ‘loud’. As for *qʔɪ́ bos* ‘cry’, its derivate *qʔɪ́botut* possesses only the meaning ‘loud’. There are instances where the participle is in fact more common than the verb. This seems to be the case of *kʔérkəbotut* ‘curly’, more common than the verb *kʔérkʔə bos* ‘go curly’, which probably only exists as a variant of *kʔúrkʔur bos* ‘eddy, swirl, babble, flow’. The case of *ɪ́rkɪ́botut* ‘thick’, derivate of *ɪ́rkɪ́ bos* ‘crowd, get squeezed, jammed’, is similar, as the adjective is derived from the compound and not from the ideophone.

Some adjectives, like *qʔánkbotu-* ‘grown up, adult (women)’ or *wáɪ́botut* ‘clear, articulate’, are obviously derived from some ideophonic compound which seems to be no longer in use.

### 5.2. Ideophonic causatives and pseudo-causatives

Only a few intransitive ideophonic verbs have a regular transitivizing (causative) construction with the auxiliary ‘do’. One of the verbs for ‘drink’ also uses this auxiliary to increase its valency by a secondary agent. Such pairs are shown in Table 23:

<i>čʔírqʔ bos</i>	‘dislocate; crack (knuckles)’	<i>čʔírqʔbos as</i>	‘make crack’ (knuckles)
<i>dím̩mus</i>	‘disperse, be destroyed’	<i>dím̩mus as</i>	‘destroy’
<i>qʔáčʔ bos</i>	‘dress up’	<i>qʔáčʔbos as</i>	‘dress up’ tr.
<i>cʔábus</i>	‘drink’	<i>cʔábus as</i>	‘make drink’
<i>háʔrš bos</i>	‘boil’ itr./tr	<i>háʔršbos as</i>	‘let boil’ tr.

Table 23. Archi causativized ideophonic compounds with regular meanings  
(data from Chumakina et al. 2007)

The examples below show the use of the two verbs ‘boil’. The underived one is ambitransitive (ex. 116 and 117) while the causative construction (ex. 118) implies an inadvertent, involuntary Agent:

(116) *ʎan*                      *kʎ'eʀrkul-li*    *haʀrʂ:ar-ʂi*    *i*.  
 water(N)(NOM)    heat-ERG    boil.IPF-CV    (N)COP  
 'The water is boiling (Russian кипит) because of the heat.'

(117) *zari*        *ʎan*                      *haʀrʂ-bo*.  
 1SG.ERG    water(N)(NOM)    boil-PF  
 'I brought some water to the boil.'

(118) *zari*    *diq'*                      *haʀrʂbo-s*    *aw*.  
 1SG.ERG soup(N)(NOM)    boil-INF    (N)do.PF  
 'I brought the soup to the boil and let it boil (too long).'

The few other verbs derived with 'do' from ideophonic verbs, as can be seen in Table 24, show unpredictable shifts in meaning:

<i>č'á'pɯs</i>	'burst, explode; shoot'	<i>č'á'pɯs as</i>	'blow up' but also 'lie, tell lies'
<i>č'é'nt' bos</i>	'squeeze, milk slowly; spit through one's teeth'	<i>č'é'nt'bos as</i>	'run over, squash'
<i>g''ár bos</i>	'hurry, walk fast'	<i>g''árbos as</i>	'eat fast'

**Table 24.** Archi causativized ideophonic compounds with irregular meanings  
 (data from Chumakina et al. 2007)

Finally, both *dárq bos* and its causative *dárqbos as* mean 'knock', while two verbs – *zít'bos as* 'slap in the face' (tr.) and *pítbos as* 'make hissing noise' (balloon, itr.) – do not have a corresponding non-auxiliarized form.

## 6. Ideophonic verb compounds in children's speech

One of the strengths of the Archi online dictionary is that it contains a large amount of 'children's vocabulary', also known as 'nursery words', i.e. words used exclusively by children or by adults when talking to them. These child speech ideophones cover the same semantic fields, but their phonotactics and phoneme inventory are restricted: the second most common phonotactic type (CVRC(V)) is absent, as are



lateral consonants and uvular occlusives, which are acquired late; and only one verb (*dá'p bos*) has a pharyngeal feature. A full list of these verbs is given in Table 25.

	PF	IPF	INF	IMPER	MSD
'bite'	<i>ʕámmus</i>	<i>ʕamár</i>	<i>ʕámmu</i>	<i>ʕámma</i>	<i>ʕámt'i</i>
'suck; ache'	<i>c'úk' bos</i>	<i>c'úk' war</i>	<i>c'úk' bo</i>	<i>c'úk' ba</i>	<i>c'úk' bumul</i>
'pee'	<i>čiš bos</i>	<i>čiš war</i>	<i>čiš bo</i>	<i>čiš ba</i>	<i>čiš't'i</i>
'walk, take steps'	<i>géd bos</i>	<i>géd war</i>	<i>géd bo</i>	<i>géd ba</i>	<i>géd bumul</i>
'fall over, strike, knock'	<i>dúk bos</i>	<i>dúk war</i>	<i>dúk bo</i>	<i>dúk ba</i>	<i>dúkt'i</i>
'fall'	<i>dá'p bos</i>	<i>dá'p war</i>	<i>dá'p bo</i>	<i>dá'p ba</i>	<i>dá'p bumul</i>
'stand'	<i>βét' bos</i>	<i>βét'er</i>	<i>βét' bo</i>	<i>βét' ba</i>	<i>βét' bumul</i>
'sit down'	<i>dák' bos</i>	<i>dák' war</i>	<i>dák' bo</i>	<i>dák' ba</i>	<i>dák'ə bumul</i>
'put on shoes'	<i>čit bos</i>	<i>čit war</i>	<i>čit bo</i>	<i>čit ba</i>	<i>čitt'i</i>
'carry on one's back'	<i>hačá bos</i>	<i>hačá war</i>	<i>hačá bo</i>	<i>hačá ba</i>	<i>hačá bumul</i>
'hit'	<i>áh bos</i>	<i>áh war</i>	<i>áh bo</i>	<i>áh ba</i>	<i>áht'i</i>
'roll'	<i>gúgo: bos</i>	<i>gúgor</i>	<i>gúgo bo</i>	<i>gúgo ba</i>	<i>gúgo bumul</i>
'sleep'	<i>lalá bos</i>	<i>lalá war</i>	<i>lalá bo</i>	<i>lalá ba</i>	(no masdar)
'toddle; leave'	<i>datá bos</i>	<i>datáwar</i>	<i>datá bo</i>	<i>datá ba</i>	<i>datá bumul</i>
'turn on (light)'	<i>t'ík'a bos</i>	<i>t'ík'a war</i>	<i>t'ík'a bo</i>	<i>t'ík'a ba</i>	<i>t'ík'a bumul</i>
'throw away'	<i>báho bos</i>	<i>báho war</i>	<i>báho bo</i>	<i>báho ba</i>	<i>bahó bumul</i>

**Table 25.** Ideophonic compounds in Archi children's speech  
(data from Chumakina et al. 2007)

The syntax of the available examples for these compounds does not show any peculiarities. Yet the absence of any gender/number agreement slot in ideophonic compounds in general makes their use particularly straightforward and probably accounts for their productivity in this semantic domain:

- (119) *laha*            *ha'tər-če-q'a-k*            *dump*            *baho + bo.*  
 child.ERG    river-OBL-INTER-LAT    ball(A)    throw + say.PF  
 'The child threw the ball in the river.'

(120) *lo*            *š:e<sup>h</sup>nt-li-t*            *dak' + bo.*  
 child(N)    chair-OBL-SUPER.ESS    sit + say.PF  
 ‘The child sat down on a chair.’

(121) *zari*    *un*            *hača + bo-qi.*  
 1.ERG    2.SG.NOM    carry + say.PF-FUT  
 ‘I will carry you on my back.’

These specialized verbs can also be used in comparisons; for instance, to denote childish behaviour:

(122) *t'i-tu-t*            *laha*            *bana*            *to-r-mi*            *gon*            *c'urar-ši*            *b-i.*  
 small-ATTR-N    child.ERG    like            DIST-F-ERG            finger(N)    suck.IPF-CV    A-COP  
 ‘She sucks her finger like a little child.’

## 7. Borrowings

A relatively large number of the coverbs used in ideophonic compounds appear to be loans from languages with which Archi has been in direct or indirect contact in the course of its history.

### 7.1. Borrowing in Archi of the Lak masdar of ‘say’

The following Lak proverbs (ex. 123 and 124) illustrate the use of ‘say’ in two Lak ideophonic compound verbs:

(123) *vajna-l*            *bivk'u-s:a*            *tuk:u-gu*            *ha?-t'-i*  
 PROX.PL.OBL-GEN    A.die.PF-ATTR            donkey-ADD            bray-say.IPF-MSD  
*b-a-ns:ar.*  
 A-do.IPF-FUT.3  
 ‘They’ll make a dead donkey bray.’

(124) *da<sup>r</sup>rβu*    *š:in-a-jn*            *uf ma-t'-ra!*  
 cold    water-OBL-SUPLAT    blow PROHIB-say.IPF-SG  
 ‘Don’t blow on cold water.’

It seems obvious that the Archi masdar ending *-t'i*, found in some but not all ideophonic compounds of this language with the same function as the native ending *-mul*, is a loan from Lak, where it belongs to the regular paradigm of the verb *t'un* 'say'. This morphological borrowing is a decisive argument in favour of an external influence on the development of ideophonic verbs in Archi.

### 7.2. Borrowed Lak ideophones in Archi

Lak uses ideophones in exactly the same manner as Archi, for the same semantic domains, and is probably at the origin of this areal feature. Archi also shares with Lak a limited number of ideophones, for instance (Table 26):

Archi		cf. Lak
<i>a'nga bos</i>	'cry' (newborn baby)	<i>o'nga t'un</i>
<i>a'nšá' bos</i>	'sneeze'	<i>a'nča t'un</i>
<i>č'ú'w bos</i>	'champ, chew with open mouth'	<i>č'a'np'a t'un</i>
<i>qélt bos</i>	'trot'	<i>χalta t'un</i>
<i>s:unt' bos</i>	'sniff'	<i>s:unt'a t'un</i>
<i>t'ánk' bos</i>	'jump'	<i>t'ank'a t'un</i>
<i>ú'nš bos</i>	'blow one's nose'	<i>o'nša t'un</i>
<i>wás bos</i>	'have doubts'	<i>was t'un</i> , cf. Arabic <i>waswasa</i>
<i>χ:árt' bos</i>	'shave'	cf. <i>χ:árt'a t'un</i> 'mow'

**Table 26.** Shared ideophonic coversbs between Archi and Lak  
(Chumakina et al. 2007, Abdullaev 2018)

### 7.3. Integration of non-ideophonic loans as ideophones

Other Archi ideophones have mostly non-ideophonic cognates in languages other than Lak. As can be seen in Table 27, these words or roots are used in Archi with a non-reduced form of the verb 'say'.

Archi	meaning	origin
<i>warč'ámi bos</i>	'greet' = 'say hello'	= Avar 'Good morning' (to a man)
<i>čúčə bos</i>	'wash (oneself or clothes)'	Avar <i>čur-</i> (without reduplication)
<i>č<sup>w</sup>árɛ<sup>s</sup> bos</i>	'flow out; pour out'	Avar <i>č<sup>w</sup>aχ-</i>
<i>čáχ<sup>s</sup> bos</i>	'flow; leak'	Avar <i>č<sup>w</sup>aχ-</i>
<i>qúrš-bos</i>	'slide, slip'	Avar <i>qúrš-</i> 'crawl'
<i>k<sup>w</sup>áš bos</i>	'ransack, rummage'	Avar <i>k<sup>w</sup>er</i> 'hand' instead of Archi <i>kul</i> (id.)
<i>k<sup>w</sup>áχ bos</i>	'grope' (in darkness)	Avar <i>k<sup>w</sup>er</i> 'hand' instead of Archi <i>kul</i> (id.)
<i>k'érs bos</i>	'nod, wag one's finger; yield'	Avar <i>kiliš:</i> 'finger'
<i>árdigi bos</i>	'forbid'	(dialectal) Avar 'don't do it!'
<i>bíz bos</i>	'appear' (grass, bristle, hair)	Turkic <i>biz</i> 'awl, prick'
<i>k'ápus</i>	'wrap up (blanket); cover' (clouds)	Turkic <i>qap-</i> 'shut'
<i>guzmús bos</i>	'mince (walk with short steps)'	<i>gəzmiş</i> , Azeri participle of <i>gəzmək</i> 'walk'
<i>gulúrči bos</i>	'fluff up' (pillow, mattress)	Azeri <i>guluɤči</i> 'handworker'
<i>lánk' bos</i>	'limp'	Persian <i>lang</i> 'lame', a regional <i>wanderwort</i>
<i>híjt bos</i>	'hug'	Arabic <i>aħat'a</i> 'surround, embrace' pres. <i>yu-ħi:t<sup>s</sup>-u</i> .
<i>wás bos</i>	'have doubts'	Arabic <i>waswasa</i> 'hesitate, to have doubts?'

**Table 27.** Loans used in Archi as ideophonic coverbs (Chumakina et al. 2007; Saidov 1967)

These findings point to the linguistic dominance of Avar, an expected and widely acknowledged fact; more contact than usually assumed between speakers of Archi (male, travelling) and those of a non-adjacent language (Azeri); and some top-down diastatic influence of Arabic as the language of religion.

## 8. Non-ideophonic compounds in *-bus* and *-mus*

The etymological origin of the second element *-mus* as an allophone of *-bus* is as certain as the fact that *-bus* is a bound variant of the free form *bos*; but of the thirty-three verbs in *-mus*, only eight have their full paradigm bearing a trace of the verb 'say' and can be confirmed to be ideophonic. These are: *é<sup>s</sup>mmus* 'cry', *č'á<sup>s</sup>mmus* 'chew', *lámmus* 'lick', *šá<sup>s</sup>mmus* 'gorge', *řámmus* 'bite' (childish), *dámmus* 'blow up', *dúmmus* 'disperse', *kámmus* 'throw'.

Other, more numerous verbs with infinitive in *-mus* and masdar in *-mul*, as well as a couple of verbs with infinitive in *-bus* and masdar in *-mul*, have no other principal part containing a trace of the verb 'to say'. Of these most probably non-ideophonic verbs, 25 have their infinitive in *-mus*. These are: *árk<sup>w</sup>mus* 'scrape, hollow out', *árt<sup>w</sup>mus* 'gnaw', *bá<sup>s</sup>kł<sup>w</sup>mus* 'press', *árχ<sup>w</sup>mus* '(get) spill(ed)', *ásmus* 'measure', *cummús* 'wash', *č'u<sup>s</sup>mmús* 'press', *c'ummús* 'sift', *dárt<sup>w</sup>mus* 'lose weight', *dá<sup>s</sup>š<sup>w</sup>mus* 'become overcast', *ékl<sup>w</sup>mus* 'perform ablution', *éq<sup>w</sup>mus* 'get a splinter; search', *éř<sup>w</sup>mus* 'sink; hide', *ét<sup>w</sup>mus*

‘tie’, *éχmus* ‘climb; hang’, *írktmus* ‘break free’, *írχ<sup>w</sup>mus* ‘work’, *kummús* ‘eat’, *kúmmus* ‘eat’ (animals), *kłummús* ‘pull’, *łummús* ‘flee’, *ókłmus* ‘flee, drive out’, *óqmus* ‘make neck and sleeves’, *óč<sup>s</sup>mus* ‘wake up’, *q<sup>w</sup>ékmus* ‘flay’, *s:ummús* ‘melt’, *χ<sup>s</sup>éχ<sup>s</sup>mus* ‘ferment’, *χ:ummús* ‘weave, knit’.

Table 28 shows the five non-ideophonic verbs in *-bus* as well as a sample of non-ideophonic verbs in *-mus* with their fourth and third gender agreeing forms:

	PF	IPF	INF	IMPER	MSD
‘boil’	<i>s:ubús</i> / <i>busbús</i>	<i>sal</i> / <i>basál</i>	<i>seté</i> / <i>besdé</i>	<i>sísi</i> / <i>bisísi</i>	<i>s:ummúl</i> / <i>busmúl</i>
‘roll over (tr), turn inside out’	<i>χ:<sup>s</sup>ubús</i> / <i>bux:<sup>s</sup>bús</i>	<i>χ:<sup>s</sup>ar</i> / <i>bax:<sup>s</sup>ár</i>	<i>χ:<sup>s</sup>eté</i> / <i>bex:<sup>s</sup>dé</i>	<i>χ:<sup>s</sup>e</i> / <i>bex:<sup>s</sup>é</i>	<i>χ:<sup>s</sup>umúl</i> / <i>bux:<sup>s</sup>múl</i>
‘take’	<i>šubús</i> / <i>bušbús</i>	<i>šur</i> / <i>bušúr</i>	<i>še</i> / <i>bešé</i>	<i>χ:a</i> / <i>bax:a</i>	<i>šummúl</i> / <i>bušmul</i>
‘paint’	<i>łubús</i> / <i>bułbús</i>	<i>łar</i> / <i>bałár</i>	<i>łeté</i> / <i>bełdé</i>	<i>łe</i> / <i>bełé</i>	<i>łumúl</i> / <i>bułmúl</i>
‘go in, enter; start doing smth’	<i>č<sup>s</sup>ubús</i> / <i>buč<sup>s</sup>bús</i>	<i>č<sup>s</sup>ur</i> / <i>buč<sup>s</sup>úr</i>	<i>č<sup>s</sup>eté</i> / <i>beždé</i>	<i>č<sup>s</sup>e</i> / <i>beč<sup>s</sup>é</i>	<i>č<sup>s</sup>ummúl</i> / <i>č<sup>s</sup>ummúl</i>
‘eat’ (human)	<i>kummús</i> / <i>bukmús</i>	<i>k<sup>w</sup>an</i> / <i>bukán</i>	<i>kunné</i> / <i>bukné</i>	<i>k<sup>w</sup>en</i> / <i>bukén</i>	<i>kummúl</i> / <i>bukmúl</i>
‘eat (animal); gorge’	<i>kúmmus</i> / <i>bukúmmus</i>	<i>kúkin</i> / <i>bukúkin</i>	<i>kunné</i> / <i>bukúnni</i>	<i>kúkin</i> / <i>bukúkin</i>	–
‘be ticklish’	<i>łorótmus</i> / <i>bołrótmus</i>	<i>łorótin</i> / <i>bołrótin</i>	<i>łorót<sup>n</sup>i</i> / <i>bołrót<sup>n</sup>i</i>	<i>łorótin</i> / <i>bołrótin</i>	<i>łorótmul</i> / <i>bołrótmul</i>
‘melt’	<i>s:ummús</i> / <i>busmús</i>	<i>s:émsin</i> / <i>bes:émsin</i>	<i>s:enné</i> / <i>besné</i>	<i>s:émsin</i> / <i>bes:émsin</i>	<i>s:émsmul</i> / <i>bes:émsmul</i>
‘rise’ (dough)	<i>χ<sup>s</sup>éχ<sup>s</sup>mus</i>	<i>χ<sup>s</sup>érχ<sup>s</sup>in</i>	<i>χ<sup>s</sup>éχ<sup>s</sup>ni</i>	<i>χ<sup>s</sup>éχ<sup>s</sup>in</i>	<i>χ<sup>s</sup>éχ<sup>s</sup>mul</i>
‘pull, stretch’	<i>kłummús</i> / <i>bukłmús</i>	<i>kłan</i> / <i>bekłén</i>	<i>kłenné</i> / <i>bekłné</i>	<i>kłen</i> / <i>bakłán</i>	<i>kłummúl</i> / <i>bekłnékul</i>
‘knead’	<i>š<sup>s</sup>umú<sup>s</sup></i> / <i>buš<sup>s</sup>mú<sup>s</sup></i>	<i>š<sup>s</sup>é<sup>s</sup>mš<sup>s</sup>in</i> / <i>beš<sup>s</sup>:é<sup>s</sup>mš<sup>s</sup>in</i>	<i>š<sup>s</sup>e<sup>s</sup>nné<sup>s</sup></i> / <i>beš<sup>s</sup>:š<sup>s</sup>né<sup>s</sup></i>	<i>š<sup>s</sup>é<sup>s</sup>mš<sup>s</sup>in</i> / <i>beš<sup>s</sup>:é<sup>s</sup>mš<sup>s</sup>in</i>	<i>š<sup>s</sup>é<sup>s</sup>mš<sup>s</sup>mul</i> / <i>beš<sup>s</sup>:é<sup>s</sup>mš<sup>s</sup>mul</i>
‘knit, weave’	<i>χ:ummús</i> / <i>buxmús</i>	<i>χ:émχ:in</i> / <i>bex:émχ:in</i>	<i>χ:enné</i> / <i>bexné</i>	<i>χ:émχ:in</i> / <i>bex:émχ:in</i>	<i>χ:émχ:mul</i> / <i>bex:émχ:mul</i>

**Table 28.** Some non-ideophonic Archi verbs with infinitives ending in *-bus* and *-mus* (data from Chumakina et al. 2007)

Most of these basic activity verbs, including the two verbs for ‘eat’ (for humans and animals respectively), have very clear cognates in other Lezgian languages and beyond. However, this topic will be left for another article.

## 9. Conclusion

Archi has a large class of more than 250 ideophones, which are a subclass of ‘coverbs’ (elements of complex verbs), used exclusively in verb compounds (and their non-inflecting derivatives), constructed with the verb ‘say’. These complex verbs, which outnumber the 163 or so simplex verbs, cover a broad range of semantic fields, including transitive action verbs, and are very common in all types of discourse. Many ideophones are semantically bleached and can be identified mainly on formal grounds because of their use with ‘say’ as a light verb.

The ubiquity of ideophonic compounds in Archi distinguishes it from the languages that are genetically closest to it: there are no such verbal compounds with the verb ‘say’ in any other Lezgian languages (except maybe Udi), where ideophones are rare in all types of speech.

Since no other member of the Lezgian branch shows such an expansive class of ideophonic coverbs, their etymology is hard to trace, but recent dictionaries will probably help identify some Archi ideophones as inherited. However, their combination with the light verb ‘say’ is a comparatively recent innovation due to the influence of Lak and is still productive, as shown by the Avar loans used as coverbs with ‘say’. It should also be noted that Avar, the most widely spoken indigenous language of Dagestan and currently the dominant language in the region where Archi is spoken, has no ideophonic compounds of the Archi, Lak or Dargwa types, and ideophones are only marginally present in its lexicon. Yet the trend set by the adoption of this construction is likely to go further because the auxiliary ‘say’ is a productive tool for creating new, inflectionally manageable verbs in a language otherwise riddled with morphological complexity.

A proper study of Lak and Dargwa ideophonic compounds remains to be undertaken, and their comparison with our findings in Archi is likely to provide more insights into the history of these languages and their contacts.

## Acknowledgements

This article could not have been written without the Archi online dictionary, from which almost all the information presented here has been extracted. Its main author, Marina Chumakina, deserves universal recognition for her outstanding work. I also wish to thank Steven Kaye, Murad Suleymanov, and the editors for English proofreading.

## Abbreviations

A	non-human animate gender	IPF	imperfective aspect
CONTLAT	‘contlative’	LAT	lative
COP	copula	M	human feminine gender
CV	converb	MSD	masdar (action noun)
DIST	distal demonstrative	N	fourth ‘inanimate’ gender
F	human feminine gender	PF	perfective aspect
FUT	future tense	PROX	proximal demonstrative
IMPER	imperative	SEQ	sequential converb
IN	locative ‘inside’	SUPER	locative ‘on’
INF	infinitive		

## References

- Abdullaev, Isa. 2018. *Laksko-russkij slovar'*. IJaLI DNC RAN, ALEF, Makhachkala.
- Arkhipov Alexandre, Daniel Michael, Khoroshkina Anna, Kibrik Aleksandr E., Kodzasov Sandro V., Olovyannikova Irina P. & Samedov Dzhilil S. 2007. Archi text corpus (1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.8247597>
- Authier, Gilles. 2009. *Grammaire kryz, langue caucasique d’Azerbaïdjan*. Louvain: Peeters.
- Chumakina, Marina, Dunstan Brown, Greville G. Corbett & Harley Quilliam. 2007. *A dictionary of Archi: Archi-Russian-English* (Online edition). University of Surrey. <http://dx.doi.org/10.15126/SMG.16/2>
- Chumakina, Marina. 2011. Morphological complexity of Archi verbs. In Gilles Authier and Timur Maisak (eds) *Tense, aspect, modality and finiteness in East Caucasian languages* (Diversitas Linguarum 30), 1-24. Bochum: Brockmeyer.

- Chumakina, Marina. 2013. Periphrasis in Archi. In Marina Chumakina & Greville G. Corbett, *Periphrasis: The role of syntax and morphology in paradigms*. (Proceedings of the British Academy 180), 27-52. Oxford: British Academy and Oxford University Press.
- Chumakina, Marina. 2015. Archi. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds), *Word formation: An international handbook of the languages of Europe* (Handbooks of Linguistics and Communication Science, HSK40). Berlin: de Gruyter Mouton.
- Chumakina, Marina. 2019. Bound but still independent: quotative and verificative in Archi. In: Matthew Baerman, Oliver Bond and Andrew Hippisley (eds). *Morphological Perspectives. Papers in Honour of Greville G. Corbett*. Edinburgh: Edinburgh University Press. 281-299.
- Chumakina, Marina & Greville G. Corbett. 2015. Gender number marking in Archi: Small is complex. In Matthew Baerman, Dunstan Brown & Greville G. Corbett (eds), *Understanding and measuring morphological complexity*, 93-116. Oxford: Oxford University Press.
- Chumakina, Marina and Bond, Oliver. 2016. Competing controllers and agreement potential. In Oliver Bond, Greville G. Corbett, Marina Chumakina and Dunstan Brown (eds), *Archi: Complexities of agreement in cross-theoretical perspective*, 77-117. Oxford Studies of Endangered Languages. Oxford: Oxford University Press.
- Ganenkov, Dmitry. 2005. *Kontaktnye lokalizacii v naxsko-dagestanskix jazykax i ix tipologičeskie paralleli*. Doctoral dissertation, State University of Moscow.
- Kibrik, Aleksandr. 1977. *Opyt strukturnogo opisanija archinskogo jazyka*. Moscow: Izdatel'stvo moskovskogo universiteta.
- Kibrik, Aleksandr. 1994. "Archi". In R. Smeets (ed.). *Indigenous languages of the Caucasus*, 297–365, Vol. 3. New York: Caravan Books.
- Saidov, Magomed Sajid. 1967. *Avarsko-russkij slovar'*. Moscow: Sovetskaya entsiklopedija

**CONTACT**

gilles.authier@gmail.com



## Appendix

## A1. Complete lists of ideophonic sensation verbs in Archi:

	PF	IPF	INF	IMPER	MSD
'smell'	<i>súnt' bos</i>	<i>sunt'ar</i>	<i>súnt' bo</i>	<i>súnt' ba</i>	<i>súntt'i</i>
'smoke; smell'	<i>p'áh bos</i>	<i>p'ahár</i>	<i>p'áh bo</i>	<i>p'áh ba</i>	<i>p'áht'i</i>
'clear up'	<i>wák' bos</i>	<i>wak'ár</i>	<i>wák' bo</i>	<i>wák' ba</i>	<i>wák't'i</i>
'break out in a rash'	<i>zímł bos</i>	<i>zímł war</i>	<i>zímł bo</i>	<i>zímł ba</i>	<i>zímł bumul</i>
'itch'	<i>χ<sup>ws</sup>ár bos</i>	<i>χ<sup>r</sup>urár</i>	<i>χ<sup>ws</sup>ár bo</i>	<i>χ<sup>ws</sup>ár ba</i>	<i>χ<sup>ws</sup>árt'i</i>
'scratch, itch'	<i>gúk:ə bos</i>	<i>guk:ór</i>	<i>gúk:ə bo</i>	<i>gúk:ə ba</i>	<i>gúk:ət'i</i>
'become cold (weather)'	<i>zámq' bos</i>	<i>zamq'ár</i>	<i>zámq' bo</i>	<i>zámq' ba</i>	<i>zámq' bumul</i>
'shrink, shiver, huddle up (cold)'	<i>q'émq'i bos</i>	<i>q'emq'ér</i>	<i>q'émq'i bo</i>	<i>q'émq'i ba</i>	<i>q'émq't'i</i>
'catch fire; get burnt' (food)	<i>pírχ bos</i>	<i>parχ:ár</i>	<i>pírχ bo</i>	<i>pírχ ba</i>	<i>pírχt'i</i>
'be overdone' (meat)	<i>túr bos</i>	<i>turár</i>	<i>túr bo</i>	<i>túr ba</i>	<i>túr bumul</i>
'become sour'	<i>q'ánc' bos</i>	<i>q'anc'ár</i>	<i>q'ánc' bo</i>	<i>q'ánc' ba</i>	<i>q'ánc't'i</i>
'ache'	<i>c'úk' bos</i>	<i>c'úk' war</i>	<i>c'úk' bo</i>	<i>c'úk' ba</i>	<i>c'úk' bumul</i>
'throb (pain)'	<i>wá'rt bos</i>	<i>wa'rtár</i>	<i>wá'rt bo</i>	<i>wá'rt ba</i>	<i>wá'rt bumul</i>
'want to eat smth'	<i>láhbos</i>	<i>lahár</i>	<i>láhbo</i>	<i>láhba</i>	<i>láht'i</i>
'be lazy, not hungry'	<i>húc' bos</i>	<i>húc' war</i>	<i>húc' bo</i>	<i>húc' ba</i>	<i>húc't'i</i>
'overripen (wheat)'	<i>k'íj bos</i>	<i>k'ijár</i>	<i>k'íj bo</i>	<i>k'íj ba</i>	<i>k'íjt'i</i>
'smoke (itr), become rich'	<i>bú'r bos</i>	<i>bú'r war</i>	<i>bú'r bo</i>	<i>bú'r ba</i>	<i>bú'rt'i</i>
'repent and reform; grow better (physically)'	<i>hínq' bos</i>	<i>hínq' war</i>	<i>hínq' bo</i>	<i>hínq' ba</i>	<i>hínq' bumul</i>
'appear' (grass, bristle, hair)	<i>bíz bos</i>	<i>bíz war</i>	<i>bíz bo</i>	<i>bíz ba</i>	<i>bíz bumul</i>

## A2. Complete lists of ideophonic weather sound verbs and animal communication verbs in Archi:

	PF	IPF	INF	IMPER	MSD
'thunder'	<i>qúbus</i>	<i>qúr</i>	<i>qúbu</i>	<i>qú ba</i>	<i>qút'i</i>
'pour' (rain)	<i>wéʒ bos</i>	<i>wéʒ war</i>	<i>wéʒ bo</i>	<i>wéʒ ba</i>	?
'drizzle'	<i>c'éc' bos</i>	<i>c'éc'er</i>	<i>c'éc' bo</i>	<i>c'éc' ba</i>	<i>c'éc'ə bumul</i>
'screech' (owl)	<i>c'ír bos</i>	<i>c'arár</i>	<i>c'ír bo</i>	<i>c'ír ba</i>	<i>c'írt'i</i>
'chirp'	<i>č'irq'ír bos</i>	<i>č'irq'ír war</i>	<i>č'irq'ír bo</i>	<i>č'irq'ír ba</i>	<i>č'irq'írt'i</i>
'buzz'	<i>zíz bos</i>	<i>zíz war</i>	<i>zíz bo</i>	<i>zíz ba</i>	<i>zíz bumul</i>
'bark'	<i>ha'mp bos</i>	<i>ha'mpár</i>	<i>há'mp bo</i>	<i>há'mp ba</i>	<i>há'mpt'i</i>
'howl'	<i>úw bos</i>	<i>uwár</i>	<i>úw bo</i>	<i>úw ba</i>	<i>úw bumul</i>
'bark'	<i>wúp bos</i>	<i>wúp war</i>	<i>wúp bo</i>	<i>wúp ba</i>	<i>wúp bumul</i>
'growl'	<i>húbus</i>	<i>hur</i>	<i>húbu</i>	<i>hú ba</i>	<i>hút'i</i>
'meow'	<i>ní'w bos</i>	<i>ní'wár</i>	<i>ní'w bo</i>	<i>ní'w ba</i>	<i>ní'wt'i</i>
'hiss, spit' (cat, snake)	<i>kʰʰ bos</i>	<i>kʰʰár</i>	<i>kʰʰ bo</i>	<i>kʰʰ ba</i>	<i>kʰʰt'i</i>
'crow' (rooster)	<i>uríʔiʔu bos</i>	<i>uríʔiʔu war</i>	<i>uríʔiʔu bo</i>	<i>uríʔiʔu ba</i>	<i>uríʔiʔu bumul</i>
'whinny'	<i>hihí bos</i>	<i>hihi wár</i>	<i>hihí bo</i>	<i>hihí ba</i>	<i>hihi bumul</i>
'sigh' (cow, horse)	<i>úh bos</i>	<i>úh war</i>	<i>úh bo</i>	<i>úh ba</i>	<i>úht'i</i>
'bray'	<i>hóʔ bos</i>	<i>haʔár</i>	<i>hóʔ bo</i>	<i>hóʔ ba</i>	<i>hóʔt'i</i>
'moo, bellow'	<i>ɓíɓ bos</i>	<i>ɓaɓár</i>	<i>ɓíɓ bo</i>	<i>ɓíɓ ba</i>	<i>ɓíɓ bumul</i>
'coo; moo'	<i>bú bos</i>	<i>búr</i>	<i>bú bo</i>	<i>bú ba</i>	<i>bú bumul</i>
'bleat' (sheep)	<i>má bos</i>	<i>má war</i>	<i>má bo</i>	<i>má ba</i>	<i>má bumul</i>
'bleat' (goat)	<i>meʔé bos</i>	<i>meʔé war</i>	<i>meʔé bo</i>	<i>meʔé ba</i>	<i>meʔé bumul</i>
'hurry sheep up'	<i>šípq'íp bos</i>	<i>šípq'íp war</i>	<i>šípq'íp bo</i>	<i>šípq'íp ba</i>	<i>šípq'íp bumul</i>
'urge sheep on'	<i>jú bos</i>	<i>júwar</i>	<i>jú bo</i>	<i>jú ba</i>	<i>jú bumul</i>

## A3. Complete lists of ideophonic intransitive movement verbs in Archi:

	PF	IPF	INF	IMPER	MSD
'fidget, swarm'	<i>ž'ál bos</i>	<i>žulár</i>	<i>ž'ál bo</i>	<i>ž'ál ba</i>	<i>ž'ált'i</i>
'crash, plump'	<i>wá'p bos</i>	<i>wá'p war</i>	<i>wá'p bo</i>	<i>wá'p ba</i>	<i>wá'p bumul</i>
'dislocate' (joint)	<i>q'árt bos</i>	<i>q'ártár</i>	<i>q'árt bo</i>	<i>q'árt ba</i>	<i>q'ártt'i</i>
'be loose' (tooth)	<i>gá'rt bos</i>	<i>gá'rtár</i>	<i>gá'rt bo</i>	<i>gá'rt ba</i>	<i>gá'rt bumul</i>
'roll out'	<i>gérgi bos</i>	<i>gérger</i>	<i>gérgi bo</i>	<i>gérgi ba</i>	<i>gérgi bumul</i>
'rock' (cot)	<i>gérkibos</i>	<i>gérker</i>	<i>gérkibo</i>	<i>gérki ba</i>	<i>gérkə bumul</i>
'go curly' (hair)	<i>k'érk'ə bos</i>	<i>k'erk'er</i>	<i>k'érk'ə bo</i>	<i>k'érk'ə ba</i>	<i>k'érk'ət'i</i>
'disperse, be destroyed'	<i>dímmus</i>	<i>damár</i>	<i>dímmu</i>	<i>dímma</i>	<i>dímt'i</i>

	PF	IPF	INF	IMPER	MSD
	PF	IPF	INF	IMPER	MSD
'be shaken, shake'	<i>dá'nk bos</i>	<i>da'nkár</i>	<i>dá'nk bo</i>	<i>dá'nk ba</i>	<i>dá'nk bumul</i>
'come apart' (hay; road)	<i>čírχ' bos</i>	<i>čírχ' war</i>	<i>čírχ' bo</i>	<i>čírχ' ba</i>	<i>čírχ' bumul</i>
'boil' (itr; tr)	<i>há'rš bos</i>	<i>ha'ršár</i>	<i>há'rš bo</i>	<i>há'rš ba</i>	<i>há'rš bumul</i>
'flow; leak'	<i>čáχ' bos</i>	<i>čaχ'ár</i>	<i>čáχ' bo</i>	<i>čáχ' ba</i>	<i>čáχ't'i</i>
'slide, slip'	<i>qúrš-bos</i>	<i>quršár</i>	<i>qúrš-bo</i>	<i>qúrš-ba</i>	<i>qúršt'i</i>
'get stuck in marshy soil'	<i>pú'rš bos</i>	<i>pu'ršár</i>	<i>pú'rš bo</i>	<i>pú'rš ba</i>	<i>pú'rš bumul</i>
'blink; start, twitch'	<i>pársbos</i>	<i>parsár</i>	<i>pársbo</i>	<i>pársba</i>	<i>párst'i</i>
'collapse'	<i>wárχ' bos</i>	<i>warχ'ár</i>	<i>wárχ' bo</i>	<i>wárχ' ba</i>	<i>wárχ't'i</i>
'crack' (wood)	<i>wírq' bos</i>	<i>wírq' war</i>	<i>wírq' bo</i>	<i>wírq' ba</i>	<i>wírq' bumul</i>
'break' (wood)	<i>žírq' bos</i>	<i>žarq'ár</i>	<i>žírq' bo</i>	<i>žírq' ba</i>	<i>žírq't'i</i>
'draw out' (smoke)	<i>c'úr bos</i>	<i>c'urár</i>	<i>c'úr bo</i>	<i>c'úr ba</i>	<i>c'úrt'i</i>
'dance'	<i>qébus</i>	<i>qer</i>	<i>qébu</i>	<i>qé ba</i>	<i>qét'i</i>
'jump'	<i>t'ánk' bos</i>	<i>t'ank'ár</i>	<i>t'ánk' bo</i>	<i>t'ánk' ba</i>	<i>t'ánk't'i</i>
'rummage'	<i>χ:<sup>w</sup>élt' bos</i>	<i>χ:<sup>w</sup>ult'ár</i>	<i>χ:<sup>w</sup>élt'bo</i>	<i>χ:<sup>w</sup>élt'ba</i>	<i>χ:<sup>w</sup>élt't'i</i>
'walk fast'	<i>dúr bos</i>	<i>durár</i>	<i>dúr bo</i>	<i>dúr ba</i>	<i>dúrt'i</i>
'mince steps'	<i>guzmús bos</i>	<i>guzmús war</i>	<i>guzmús bo</i>	<i>guzmús ba</i>	<i>guzmús bumul</i>
'speed up'	<i>t'ír'k' bos</i>	<i>t'ark'ár</i>	<i>t'ír'k' bo</i>	<i>t'ír'k' ba</i>	<i>t'ír'k't'i</i>
'trot' (horses)	<i>qélt bos</i>	<i>qéltar</i>	<i>qélt bo</i>	<i>qélt ba</i>	<i>qéltt'i</i>
'stroll, walk'	<i>qéc-šec bos</i>	<i>qéc-šec war</i>	<i>qéc-šec bo</i>	<i>qéc-šec ba</i>	<i>qéc-šec bumul</i>
'walk, run' (cars)	<i>qéc bos</i>	<i>qécer</i>	<i>qéc bo</i>	<i>qéc ba</i>	<i>qéc bumul</i>
'crawl'	<i>ɬúrχ' bos</i>	<i>ɬurχ'ár</i>	<i>ɬúrχ' bo</i>	<i>ɬúrχ' ba</i>	<i>ɬúrχ't'i</i>
'shuffle one's feet'	<i>ɬérχ' bos</i>	<i>ɬurχ'ár</i>	<i>ɬérχ' bo</i>	<i>ɬérχ' ba</i>	<i>ɬurχ't'i</i>
'fly'	<i>párχ bos</i>	<i>parχár</i>	<i>párχ bo</i>	<i>párχ ba</i>	<i>párχt'i</i>
'stream, fly'	<i>pár bos</i>	<i>parár</i>	<i>pár bo</i>	<i>pár ba</i>	<i>párt'i</i>
'go fast' (livestock)	<i>qút bos</i>	<i>qút war</i>	<i>qút bo</i>	<i>qút ba</i>	<i>qút bumul</i>
'limp'	<i>lánk' bos</i>	<i>lank'ár</i>	<i>lánk' bo</i>	<i>lánk'ba</i>	<i>lánk't'i</i>
'wag one's finger; yield'	<i>k'érsš bos</i>	<i>k'érsš:er</i>	<i>k'érsš bo</i>	<i>k'érsš ba</i>	<i>k'érsšt'i</i>

## A4. Complete lists of ideophonic transitive movement verbs in Archi:

	PF	IPF	INF	IMPER	MSD
'squeeze, milk slowly, spit through one's teeth'	<i>č'é<sup>s</sup>nt' bos</i>	<i>č'é<sup>s</sup>nt'er</i>	<i>č'é<sup>s</sup>nt' bo</i>	<i>č'é<sup>s</sup>nt' ba</i>	<i>č'é<sup>s</sup>ntt'i</i>
'throw; wave'	<i>láp:us</i>	<i>lap'ár</i>	<i>láp:u</i>	<i>láp:a</i>	<i>láp:ukul</i>
'rock vigorously (baby); eat one's fill'	<i>bú<sup>s</sup>rt' bos</i>	<i>bu<sup>s</sup>rt'ár</i>	<i>bú<sup>s</sup>rt' bo</i>	<i>bú<sup>s</sup>rt' ba</i>	<i>bú<sup>s</sup>rtt'i</i>
'pull' (fleece), 'fluff up' (pillow, feather bed)	<i>gulúrči bos</i>	<i>gulúrčor</i>	<i>gulúrči bo</i>	<i>gulúrči ba</i>	<i>gulúrčit'i</i>
'limp; adjust' (eg., a pack on one's back)	<i>húrk' bos</i>	<i>hurk'ár</i>	<i>húrk' bo</i>	<i>húrk' ba</i>	<i>húrk'ə bumul</i>
'incite, stir up against'	<i>husú bos</i>	<i>husú war</i>	<i>husú bo</i>	<i>husú ba</i>	<i>husú bumul</i>
'slam' (door)	<i>q<sup>w</sup>ank' bos</i>	<i>q<sup>w</sup>ank' war</i>	<i>q<sup>w</sup>ank' bo</i>	<i>q<sup>w</sup>ank' ba</i>	<i>q<sup>w</sup>ank' t'i</i>
'push, slap lightly, hammer' (lightly)	<i>q'úc' bos</i>	<i>q'úc' war</i>	<i>q'úc' bo</i>	<i>q'úc' ba</i>	<i>q'úct'i</i>
'separate lambs'	<i>deré bos</i>	<i>dere wár</i>	<i>deré bo</i>	<i>deré ba</i>	<i>deré bummul</i>
'ransack, rummage'	<i>k<sup>w</sup>áš bos</i>	<i>kušár</i>	<i>k<sup>w</sup>áš bo</i>	<i>k<sup>w</sup>áš ba</i>	<i>k<sup>w</sup>áš bumul</i>
'hug'	<i>híjt bos</i>	<i>hijtár</i>	<i>híjt bo</i>	<i>híjt ba</i>	<i>híjtt'i</i>
'snatch'	<i>χáp:us</i>	<i>χapár</i>	<i>χáp:u</i>	<i>χáp:a</i>	<i>χápt'i</i>
'tap; peck; do fast'	<i>k'út' bos</i>	<i>k'út'ár</i>	<i>k'út' bo</i>	<i>k'út' ba</i>	<i>k'útt'i</i>
'caress' (a child)	<i>dítan bos</i>	<i>dítan war</i>	<i>dítan bo</i>	<i>dítan ba</i>	<i>dítan bumul</i>
'turn back, drive cattle; ruffle'	<i>dáɓ bos</i>	<i>daɓár</i>	<i>dáɓ bo</i>	<i>dáɓ ba</i>	<i>dáɓt'i</i>
'flow out; pour out'	<i>č<sup>w</sup>árɓ<sup>s</sup> bos</i>	<i>čurɓ<sup>s</sup>ár</i>	<i>č<sup>w</sup>árɓ<sup>s</sup> bo</i>	<i>č<sup>w</sup>árɓ<sup>s</sup> ba</i>	<i>č<sup>w</sup>árɓ<sup>s</sup>t'i</i>
'turn, squeeze'	<i>q<sup>w</sup>érq'ə bos</i>	<i>q<sup>w</sup>érq'er</i>	<i>q<sup>w</sup>érq'əbo</i>	<i>q<sup>w</sup>érq'ə ba</i>	<i>q<sup>w</sup>érq'ət'i</i>
	PF	IPF	INF	IMPER	MSD
'hold smb/smith with an iron hand'	<i>q<sup>w</sup>árɩ bos</i>	<i>q'urɩár</i>	<i>q<sup>w</sup>árɩ bo</i>	<i>q<sup>w</sup>árɩ ba</i>	<i>q<sup>w</sup>árɩt'i</i>

	PF	IPF	INF	IMPER	MSD
'push, make fall'	<i>χít' bos</i>	<i>χat'ár</i>	<i>χít' bo</i>	<i>χít' ba</i>	<i>χít't'i</i>
'stuff, fill'	<i>zánk bos</i>	<i>zánk war</i>	<i>zánk bo</i>	<i>zánk ba</i>	<i>zánkt'i</i>
'butt; stick in, thrust'	<i>q'ébus</i>	<i>q'ér</i>	<i>q'ébu</i>	<i>q'é-ba</i>	<i>q'é't'i</i>
'jam, crush, squash'	<i>úrkt' bos</i>	<i>urkt'ár</i>	<i>úrkt' bo</i>	<i>úrkt' ba</i>	<i>úrkt't'i</i>
'push, cram, stuff in'	<i>š<sup>w</sup>élš'i bos</i>	<i>š<sup>w</sup>elš'ér</i>	<i>š<sup>w</sup>élš'i bo</i>	<i>š<sup>w</sup>élš'i ba</i>	<i>š<sup>w</sup>élš'i bumul</i>
'coat, smear'	<i>súsie bos</i>	<i>sus:ór</i>	<i>súsie bo</i>	<i>súsie ba</i>	<i>sús:et'i</i>
'dig, take off a layer'	<i>qért' bos</i>	<i>qért'er</i>	<i>qért' bo</i>	<i>qért' ba</i>	<i>qértt'i</i>
'hit, butt' (calf)	<i>q'úč bos</i>	<i>q'úč war</i>	<i>q'úč bo</i>	<i>q'úč ba</i>	<i>q'úč bumul</i>
'wrap' (blanket); 'cover' (clouds)	<i>k'á'p:us</i>	<i>k'a'p:ár</i>	<i>k'á'p:u</i>	<i>k'á'p:a</i>	<i>k'á'pt'i</i>
'fill to the brim'	<i>zérzi bos</i>	<i>zérzer</i>	<i>zérzi bo</i>	<i>zérzi ba</i>	<i>zérzət'i</i>
'sprinkle salt'	<i>zúz bos</i>	<i>zúz war</i>	<i>zúz bo</i>	<i>zúz ba</i>	<i>zúz bumul</i>
'shake' (clothes)	<i>q'áp:us</i>	<i>q'ap:ár</i>	<i>q'áp:u</i>	<i>q'áp:a</i>	<i>q'ápt'i</i>
'push (people), stick a needle; hammer'	<i>q'éč bos</i>	<i>q'éčer</i>	<i>q'éč bo</i>	<i>q'éč ba</i>	<i>q'éčt'i</i>
'rub; rub oneself'	<i>kłúr bos</i>	<i>kłurár</i>	<i>kłúr bo</i>	<i>kłúr ba</i>	<i>kłúrt'i</i>
'splash, spatter with'	<i>tért bos</i>	<i>tertér</i>	<i>tért bo</i>	<i>tért ba</i>	<i>tért bumul</i>
'throw; shoot' (gun)	<i>kámmus</i>	<i>kamár</i>	<i>kámmu</i>	<i>kámma</i>	<i>kámt'i</i>

## A5. Complete lists of ideophonic coverbs with CVC template:

<i>bá? bos</i>	'kiss'	<i>c'úr bos</i>	'draw out' (smoke)
<i>bíz bos</i>	'appear' (grass, bristle, hair)	<i>č'ú<sup>w</sup> bos</i>	'champ, chew with open mouth'
<i>bú'r bos</i>	'smoke (itr), become rich'	<i>čák' bos</i>	'lick clean' (plates)
<i>č'ák' bos</i>	'chatter'	<i>čáχ<sup>s</sup> bos</i>	'flow; leak'
<i>č'á<sup>m</sup>mmus</i>	'chew'	<i>čřł bos</i>	'sizzle'
<i>č'á'p:us</i>	'burst, explode; crack; shoot'	<i>dáb bos</i>	'turn back; drive cattle; ruffle'
<i>c'íq' bos</i>	'fart'	<i>dé'j bos</i>	'resound' (echo)
<i>c'ír bos</i>	'screech' (owl)	<i>dámmus</i>	'explode'
<i>c'úk' bos</i>	'ache'		

<i>dímmus</i>	‘disperse, be destroyed’	<i>qat’ bos</i>	‘measure one’s strength with smb,’
<i>dúr bos</i>	‘walk fast’	<i>qéc bos</i>	‘walk, run’ (cars)
<i>dú’t’ bos</i>	‘blow (trumpet)’	<i>q’éc’ bos</i>	‘push (people), stick a needle; hammer’
<i>g<sup>w</sup>ár bos as</i>	‘eat sth fast’	<i>qíp:us</i>	‘defecate’
<i>há’h bos</i>	‘breathe’	<i>qúc bos</i>	‘sew with short stitches’
<i>hákt’ bo</i>	‘yawn’	<i>q’úč bos</i>	‘hit, butt (calf)’
<i>hát bos</i>	‘scold’	<i>qút bos</i>	‘go fast (livestock)’
<i>hīj bos</i>	‘plough’	<i>κ’áz bos</i>	‘cut’ (grass)
<i>hó’? bos</i>	‘bray’	<i>κ’úž bos</i>	‘rub hard’ (clothes)
<i>húbus</i>	‘growl’	<i>κ<sup>ws</sup>ár bos</i>	‘crack (voice)’
<i>húč’ bos</i>	‘be lazy, not hungry’	<i>š<sup>w</sup>ít’ bos</i>	‘whistle’
<i>k’á’p:us</i>	‘wrap’ (blanket); ‘cover’ (clouds)	<i>šá’mmus</i>	‘gorge’
<i>k’fj bos</i>	‘overripen (wheat)’	<i>šá’mmus</i>	‘gorge’
<i>k’úr bos</i>	‘take a nap’	<i>sém bos</i>	‘eat too little and unwillingly’
<i>k’út’ bos</i>	‘tap; peck; do smth fast’	<i>šúš bos</i>	to whisper’
<i>kámmus</i>	‘throw; shoot (gun)’	<i>túr bos</i>	‘be overdone, be spoilt’ (meat)
<i>křř bos</i>	‘hiss, spit (cat, snake)’	<i>wák’ bos</i>	‘clear up’
<i>křúr bos</i>	‘rub; rub oneself’	<i>wá’p’ bos</i>	‘crash, plump’
<i>k<sup>w</sup>áš bos</i>	‘ransack, rummage’	<i>wás bos</i>	‘have doubts’
<i>láhbos</i>	‘want to eat smth’	<i>wúp bos</i>	‘bark’
<i>lammus</i>	‘lick’	<i>záp:os</i>	‘invite’
<i>láp:us</i>	‘throw; wave’	<i>zúz bos</i>	‘sprinkle salt’
<i>p’áh bos</i>	‘smoke; smell’	<i>ž<sup>w</sup>ál bos</i>	‘fidget, swarm’
<i>p’á? bos</i>	‘kiss’	<i>ζóp’ bos</i>	‘burp’
<i>pár bos</i>	‘stream, fly’	<i>χáp:us</i>	‘snatch’
<i>pút bos</i>	‘stew (nettle); sputter’	<i>χάq<sup>s</sup> bos</i>	‘throw up’
<i>q’áč’ bos</i>	‘dress up’	<i>χít’ bos</i>	‘push, make fall’
<i>q’áp:us</i>	‘shake (washed clothes)’	<i>χ<sup>ws</sup>ár bos</i>	‘itch’
<i>q’éc’ bos</i>	‘nag, bother’	<i>χ<sup>w</sup>ít’ bós</i>	‘comb threads while weaving’
<i>q’ix bos</i>	‘yell’		
<i>q’úc’ bos</i>	‘push, slap lightly, hammer’		
<i>q<sup>ws</sup>ár- bos</i>	‘rumble’ (stomach)		

## A6. Complete lists of ideophonic coverbs with CVRC template:

<i>bú<sup>rt</sup> bos</i>	‘rock’ (baby); ‘eat one’s fill’	<i>qélt bos</i>	‘trot’ (horses)
<i>č<sup>é</sup>nt<sup>t</sup> bos</i>	‘squeeze, milk slowly; spit through one’s teeth’	<i>qénz bos</i>	‘spin thread, wool’
<i>č<sup>í</sup>rq<sup>’</sup> bos</i>	‘dislocate; crack’ (knuckles)	<i>qért<sup>’</sup> bos</i>	‘dig, take off a layer of earth’
<i>čír<sup>χ</sup> bos</i>	‘come apart’ (hay; road)	<i>qúrš-bos</i>	‘slide, slip’
<i>č<sup>w</sup>ár<sup>β</sup> bos</i>	‘flow out; pour out’	<i>βúrkl<sup>’</sup> bos</i>	‘jam, crush, squash’
<i>č<sup>w</sup>ár<sup>χ</sup> bos</i>	‘stamp, clap’	<i>sínk<sup>’</sup> bos</i>	‘sob’
<i>dá<sup>nk</sup> bos</i>	‘pulse’ (heart)	<i>t<sup>’</sup>ánk<sup>’</sup> bos</i>	‘jump’
<i>dá<sup>nk</sup> bos</i>	‘be shaken, shake’	<i>t<sup>’</sup>rk<sup>’</sup> bos</i>	‘speed up’
<i>dárq bos</i>	‘knock’	<i>tár<sup>χ</sup> bos</i>	‘twist wool’
<i>gá<sup>rt</sup>bos</i>	‘be loose’ (tooth)	<i>tért bos</i>	‘splash, spatter with’
<i>ha<sup>mp</sup> bos</i>	‘bark’	<i>wár<sup>χ</sup> bos</i>	‘collapse’
<i>há<sup>r</sup>š bos</i>	‘boil’ (itr; tr)	<i>wé<sup>l</sup> bos</i>	‘squelch’ (water in boots)
<i>hérq<sup>s</sup> bos</i>	‘belch; hiccough’	<i>wírq<sup>’</sup> bos</i>	‘crack (wood)’
<i>húrk<sup>’</sup> bos</i>	‘limp; adjust’ (pack)	<i>zánk bos</i>	‘stuff, fill’
<i>k<sup>ent</sup> bos</i>	‘drip’	<i>žírq bos</i>	‘brag’
<i>k<sup>ér</sup>š bos</i>	‘wag one’s finger; yield’	<i>žírq<sup>’</sup> bos</i>	‘break’ (wood)
<i>łáls bos</i>	‘stammer’	<i>χ:ánk<sup>’</sup> bos</i>	‘snore’
<i>łúr<sup>χ</sup> bos</i>	‘crawl’	<i>χ:árt<sup>’</sup> bos</i>	‘shave’
<i>lánk<sup>’</sup> bos</i>	‘limp’	<i>χ:<sup>w</sup>élt<sup>’</sup> bos</i>	‘rummage’
<i>łér<sup>χ</sup> bos</i>	‘shuffle one’s feet’ (people)	<i>a<sup>s</sup>ngá<sup>s</sup> bos</i>	‘cry’ (baby)
<i>ní<sup>w</sup>bos</i>	‘meow’	<i>a<sup>s</sup>nšá<sup>s</sup> bos</i>	‘sneeze’
<i>pálc<sup>’</sup> bos</i>	‘hew, grind smooth’	<i>β<sup>é</sup>m<sup>β</sup>ə bos</i>	‘gnaw’ (dogs)
<i>pársbos</i>	‘blink; start, twitch’	<i>t<sup>é</sup>nt<sup>ə</sup> bos</i>	‘overeat (animals), glut’
<i>pár<sup>χ</sup> bos</i>	‘fly’	<i>gúr<sup>q</sup>s bos</i>	‘swallow without chewing’
<i>pír<sup>q</sup> bos</i>	‘split, crack (wood, dry skin)’	<i>há<sup>n</sup>č<sup>’</sup> bos</i>	‘graze’
<i>pír<sup>χ</sup> bos</i>	‘startle’ (heart)	<i>χ:úrk<sup>’</sup> bos</i>	‘slurp’
<i>pú<sup>r</sup>š bos</i>	‘get stuck in marshy soil’	<i>súnt<sup>’</sup> bos</i>	‘smell’
<i>q<sup>ár</sup>š bos</i>	‘gnash, grit’ (teeth)	<i>zímł bos</i>	‘break out in a rash’
<i>q<sup>s</sup>árt<sup>’</sup> bos</i>	‘dislocate’ (joint)	<i>zámq<sup>’</sup> bos</i>	‘become very cold’ (weather)
<i>q<sup>é</sup>nc<sup>’</sup> bos</i>	‘shear sheep, trim hair’	<i>q<sup>é</sup>mq<sup>’</sup>i bos</i>	‘huddle up’ (from cold)
<i>q<sup>s</sup>ént<sup>’</sup> bos</i>	‘pinch’	<i>pír<sup>χ</sup> bos</i>	‘take fire; get burnt’ (food)
<i>q<sup>w</sup>ank<sup>’</sup> bos</i>	‘slam’ (door)	<i>q<sup>á</sup>nc<sup>’</sup> bos</i>	‘become sour’
<i>q<sup>w</sup>ánt<sup>’</sup> bos</i>	‘say, order sharply’	<i>wá<sup>rt</sup> bos</i>	‘throb’ (pain)
<i>q<sup>w</sup>árt<sup>’</sup> bos</i>	‘hold with an iron hand’	<i>hínq<sup>’</sup> bos</i>	‘repent and reform’
		<i>híjt bos</i>	‘hug’

## A7. Complete lists of ideophonic coverbs with other templates:

CV	<i>q'ėbus</i>	'butt; stick in, thrust'	<i>á'?</i>	<i>bos</i>	'to call, tell, ask, invite'
	<i>qėbus</i>	'dance'	<i>ė'mmus</i>		'cry'
	<i>bá'bus</i>	'talk'	<i>úw bos</i>		'howl'
	<i>χábus</i>	'sing' / 'play'	VCV	<i>út'u bos</i>	'refuse'
	<i>húbus</i>	'growl'		<i>uńú bos</i>	'cough'
	<i>rá' bos</i>	'croak; weep' (ironic)	CVCV	<i>hačá bos</i>	'carry on one's back'
	<i>húbus</i>	'blow' (wind or person)		<i>deré bos</i>	'separate lambs'
	<i>tú bus</i>	'spit'		<i>hus:ú bos</i>	'set on, incite, stir up'
	<i>χ:úbus</i>	'sip smth hot'	VRC	<i>ú'nš bos</i>	'blow one's nose'
	<i>c'ábus</i>	'drink (cold)'	CVCVR	<i>đítan bos</i>	'caress' (a child)
VC	<i>áč bos</i>	'clear throat and spit'			

## A8. Complete lists of ideophonic coverbs with reduplication:

CVC	<i>c'ėc' bos</i>	'drizzle'
	<i>zız bos</i>	'buzz'
	<i>bıı bos</i>	'moo, bellow'
CVCV	<i>čúčə bos</i>	'wash' (oneself or clothes)
	<i>hıhı bos</i>	'whinny'
	<i>gúkə bos</i>	'scratch, to itch'
	<i>č'w'ėč'ə bos</i>	'nibble'
	<i>χ:áč:u bos</i>	'eat food made of <i>dac'on</i> '
	<i>χ:úχ:u bos</i>	'saw'
	<i>sése bos</i>	'roast grain'
	<i>súsie bos</i>	'coat, smear'
CVRCV	<i>t'ėnt'ə bos</i>	'overeat (animals), glut (people)'
	<i>gėrgi bos</i>	'roll out'
	<i>gėrkııbos</i>	'rock' (cot)
	<i>k'ėrk'ə bos</i>	'go curly' (hair)
	<i>q'w'ėrq'ə bos</i>	'turn (smth), squeeze'
	<i>š:w'ėłš:i bos</i>	'push, cram, stuff in'
	<i>β'ėmβ'ə bos</i>	'gnaw' (dogs)
	<i>β'ėrβə bos</i>	'break' (crockery)
	<i>zėrzi bos</i>	'fill to the brim'
CVRCVR	<i>k'úrk'ur bos</i>	'purl, babble, flow' (spring); 'fall asleep'



	<i>č'irq'ir</i>	'chirp'
	<i>dúmɤum bos</i>	'mumble, grumble'
	<i>búrɤ'ur bos</i>	'grumble, mumble'
CVC-CVC	<i>č'á'p'-q'ap' bos</i>	'exchange fire, trade shots' cf. <i>č'á'pus</i> 'burst, explode; shoot'
	<i>š'ipq'ip bos</i>	'hurry sheep up'
	<i>š:'wít'q<sup>w</sup>'it' bos</i>	'whistle softly' (melody)
	<i>qécšec bos</i>	'stroll, walk'
	<i>šúšq'us bos</i>	'gossip'
	<i>há'hčit bos</i>	'pant, puff'
other	<i>urí'ʔiʔu<sup>s</sup> bos</i>	'crow' (rooster)

# Ideophones in Upper Guinea Creoles: a comparative study

NICOLAS QUINT<sup>1</sup>, NOËL BERNARD BIAGUI<sup>2</sup>

<sup>1</sup>CNRS, UMR 8135 LANGAGE, LANGUES ET CULTURES D'AFRIQUE

<sup>2</sup>CENTRE DE LINGUISTIQUE APPLIQUEE DE DAKAR (UCAD)

Submitted: 02/04/2023 Revised version: 01/03/2024

Accepted: 04/03/2024 Published: 12/06/2024



Articles are published under a Creative Commons Attribution 4.0 International License (The authors remain the copyright holders and grant third parties the right to use, reproduce, and share the article).

## Abstract

The Upper Guinea Creoles (UGCs) are a family of closely related Afro-Portuguese languages, comprising three branches: continental (Casamance and Guinea-Bissau), insular (Cape Verde) and ABC (Dutch Antilles). Several continental and insular UGC varieties make use of a specific set of adverbs which can be called “ideophones” following Dingemanse’s (2012) definition of the term. This paper aims at providing a comprehensive study of UGC ideophones. Based on a database collected from native speakers, it characterizes the main phonological, morphosyntactic and semantic features of ideophonic items in UGCs. In addition, it investigates the origins of UGC ideophones and compares the use and behavior of this word class in continental and insular UGCs, showing the degree of both African and Portuguese influence on each UGC variety considered.

**Keywords:** Afro-Portuguese, ideophones, language contact, Niger-Congo, Upper Guinea Creoles.

## 1. Introduction

Like many Niger-Congo languages, all African members of the Upper Guinea Creoles (UGCs) family, a group of Afro-Portuguese languages spoken in West Africa and in three Dutch Caribbean islands, make use of a specific set of adverbs which can be called “ideophones” (IDEO) following Dingemanse’s (2012) definition of the term. Examples (1)-(3), taken from Casamance Creole (CC), illustrate one particular

subcategory of ideophones, namely verb intensifiers, which are highly specialized and each intensify only one or a few verb(s) of the language in question:<sup>1</sup>

- (1) CC *bran̄ku*<sup>2</sup> ‘be white’ + *fandan̄* (IDEO) > *bran̄ku fandan̄* ‘be **very** white, be **as white as snow**’
- (2) CC *pretu* ‘be black’ + *nok* (IDEO) > *pretu nok* ‘be **very** black, be **pitch-black**, be **as black as coal**’
- (3) CC *kalá* ‘be/keep silent’ + *mik* (IDEO) > *kalá mik* ‘be **very** silent, be **as quiet as a mouse**, be **dead silent**’

The goal of this study is threefold: first, it aims at giving a detailed description of the main phonological, morphosyntactic and semantic characteristics of UGC ideophones; second, it takes stock of the question of their origin; and third, it provides a comparative perspective across two of the three branches of UGCs: the continental branch (Casamance and Guinea-Bissau varieties) and the insular branch (Capeverdean varieties).

The structure of the paper is as follows: in Section 2, we provide the reader with an introduction to the UGC family. Then, in Section 3, we give a brief overview of UGC ideophones, against the background of the typological literature. We explain why our corpus is skewed in favor of verb-intensifying ideophones and mention relevant publications on UGC ideophones. Section 4 provides details about the data and our methodology. Section 5 deals with continental UGC ideophones, whereas Section 6 examines insular UGC ideophones and compares them with their continental counterparts. Section 7 summarizes the main results of this study.

---

<sup>1</sup> For the intensification of non-verbal elements, see §5.2.2.

<sup>2</sup> Unless otherwise specified, UGC data are transcribed orthographically with phonologically-based conventions that generally follow the rules used by Biagui in his grammar (2018: 31-41) for continental UGC data (Casamance and Guinea-Bissau) and the prescriptions of the ALUPEC (*Alfabeto Unificado para a Escrita do Caboverdiano* = Unified Alphabet for Capeverdean writing) for Capeverdean data. Data drawn from other authors were adapted to these rules. Note that in ALUPEC, the digraphs {dj}, {nh} and {tx} code for one phoneme each, namely /dʒ/~ʒ/, /ɲ/ and /tʃ/~ç/, respectively.

## **2. Upper Guinea Creoles: some basic facts**

Apart from the specific references mentioned throughout this section, most of the information presented hereafter is based on Biagui et al. (forthcoming), Quint & Moreira Tavares (2019), Jacobs (2012), and Quint (2000b).

### ***2.1. Main subgroupings and varieties***

The Upper Guinea Creoles are a group of phylogenetically related Afro-Portuguese languages traditionally spoken in West Africa and in the Dutch Antilles. They comprise three main branches:

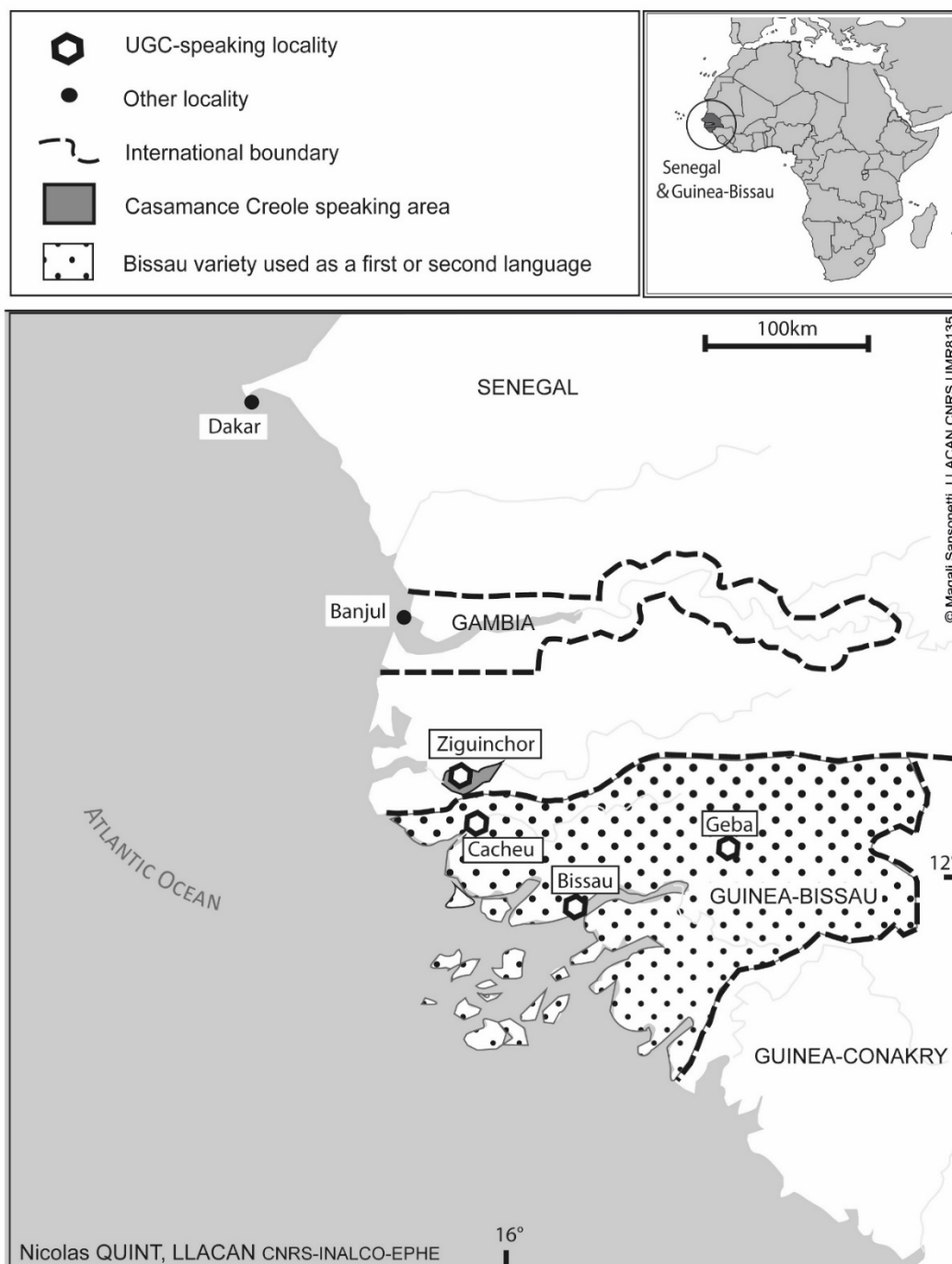
(i) Continental UGCs (pov), spoken in Guinea-Bissau and Senegal (Map 1): in Senegal, only one UGC variety is spoken today, namely Casamance Creole, traditionally used in the city of Ziguinchor and its surroundings (in the region of Lower Casamance). It has approximately 20.000 speakers, 10.000 of whom are native.

In Guinea-Bissau, at least three distinct UGC varieties are recorded: the Bissau variety, originally spoken in the capital city of Guinea-Bissau, has now become the main vehicular language of the whole country. It has nearly 2.000.000 speakers, over 500.000 of whom are native. The Cacheu and Geba varieties, spoken in the eponymous cities situated in Northwestern and Eastern Guinea-Bissau, respectively, are currently on the verge of extinction (as they are being replaced by the Bissau variety), and each of them is used or remembered by only a handful of elderly people (Quint 2023: 452). The Cacheu variety has been shown to be historically related to Casamance Creole (Quint 2023: 451, 453; Quint & Moreira Tavares 2019: 135-139; Biagui 2018: 18) whereas the Geba variety clusters with the Bissau variety.

(ii) Insular (or Capeverdean) UGCs (Map 2) are spoken by 1.000.000 native speakers, 500.000 of whom live in the archipelago of Cape Verde and the rest in diasporic communities, mainly in the US, Portugal, France, Netherlands and Senegal (Quint 2009b). Insular UGCs can be further split into two sub-branches: Sotavento ('Leeward') Capeverdean is spoken in the four southern islands of Brava, Fogo, Maio, and Santiago. Barlavento ('Windward') Capeverdean is spoken in the five northern islands of Boa Vista, Sal, Sant'Antão, São Nicolau, and São Vicente.

(iii) Papiamentu (or ABC UGC, (pap)) is spoken in the three Caribbean islands of Aruba, Bonaire and Curaçao (the ABC islands) by at least 350.000 people including a

diaspora of several tens of thousands of people in the Netherlands. Each island has its own variety.



**Map 1:** Continental Upper Guinea Creoles.



**Map 2:** Insular (Capeverdean) Upper Guinea Creoles.

The degree of mutual understanding between the different UGCs is variable: speakers of any two continental varieties can easily chat with each other without any previous contact, and the same is true for speakers of the various ABC varieties. The insular branch shows a higher degree of internal differentiation, in particular between Barlavento and Sotavento: users of rural varieties of each of these subbranches may find it difficult to fully understand each other. Between speakers of different UGC branches, a basic understanding is always possible due to the structural and lexical features shared by all UGCs. However, the linguistic distance between each of the three branches is such that speakers must usually actively learn some elements of each other’s varieties before being able to communicate effectively. In other words,

the UGCs can be considered as a group of closely related languages, and each branch of the family as a dialect cluster.

## ***2.2. African and Portuguese origins***

All UGCs are characterized by a common Afro-Portuguese lexical core, including a Portuguese component and an African substrate.

Except for contemporary Papiamentu, over 80% of the core lexicon (e.g. Swadesh list) of all UGCs is of Portuguese origin, and therefore UGCs can be described as Portuguese-based Creole languages. Note that this Portuguese component is not derived from today's Portuguese but from classical Portuguese, i.e. the stage of the language that was spoken between 1450 and 1550, at the time when UGCs first developed (see below).

The African substrate is, essentially, derived from the three following languages, in decreasing order of importance: Mandinka (mnk), Wolof (wof/wol) and Temne (tem) (Biagui et al. forthcoming; Quint & Moreira Tavares 2019: 124-126; Rougé 1999). These three languages all belong to the Niger-Congo phylum. Wolof and Temne belong to the Atlantic branch of this phylum while Mandinka belongs to the Mande branch.

It is the existence of this Afro-Portuguese lexical core, combined with a series of striking phonological and structural similarities, which has led scholars to posit the existence of proto-UGC, an ancestor common to all UGCs, which was different from Portuguese.<sup>3</sup> Proto-UGC must have developed during the second half of the 15<sup>th</sup> century and was in all likelihood fully shaped from 1550 onwards. At the beginning, proto-UGC may have been used as a pidgin (i.e. a non-native contact variety) allowing exchanges between Portuguese sailors and West African people. At any rate, it soon became the native language of some Afro-Portuguese communities, initially in Cape Verde, on the island of Santiago, the first Capeverdean island to have been settled by the Portuguese (ca. 1460) and therefore the most plausible cradle for all UGCs. From then on, UGC varieties developed on the island of Fogo (settled between 1480 and 1490, see Moreira Tavares (2020: 24-25)) and then progressively on the remaining islands of the archipelago of Cape Verde.

Continental varieties must have developed in the second half of the 16<sup>th</sup> century (Biagui et al. forthcoming), from an offshoot of the incipient Capeverdean (insular) UGC. As all continental UGCs are obviously closely related and share features not

---

<sup>3</sup> For a first tentative reconstruction of proto-UGC lexicon and grammar, see Quint (2000b: 197-208, 307-318).

found in other UGCs, it is reasonable to posit the existence of a common ancestor for these varieties, namely proto-continental UGC, which must have appeared in Cacheu, the first Portuguese permanent settlement on West African shores.

Papiamentu (or ABC) UGC emerged in the first half of the 17<sup>th</sup> century, when the Dutch colonized Curaçao and then the two other islands (Aruba and Bonaire) where it is still spoken.

UGCs are mostly analytic languages, and make a much more limited use of bound morphology than Portuguese (and more generally Romance languages). They all follow a strict SVO, determiner + noun, noun + genitive word order.

### 3. Ideophones in UGCs: a brief overview

#### 3.1. Defining the category of ideophones in UGCs

In UGCs, as in many languages of the world, ideophones are a subtype of adverbs that typically combine with a verb and depict different semantic nuances associated with this verb, such as intensification (1-3), immediacy (4) or different sensory modalities (see Dingemanse 2012: 663 for more details), such as sounds (in which case the ideophones may also be called “onomatopoeias”<sup>4</sup>) (5) or visual patterns (6).

#### (4) Cacheu (Guinea-Bissau) Creole

(a) *labantá* ‘stand up’ + *fakat* (IDEO) > *labantá fakat* ‘stand up **at once**’

(b) *sai* ‘get out’ + *fut* (IDEO) > *sai fut* ‘get out **at once/quickly**’

#### (5) Casamance Creole (CC)

(a) *kay* ‘fall’ + *furbap* (IDEO) > *kay furbap* ‘fall **producing the typical noise of something sinking into rice bran**’

(b) *ñemé*<sup>5</sup> ‘chew’ + *cákum-cákum* (IDEO) > *ñemé cákum-cákum* ‘chew **noisily**’

---

<sup>4</sup> The concepts of sound ideophone and onomatopoeia do not always overlap in all languages. Within the scope of this paper, we will consider that all UGC sound ideophones can also be considered as onomatopoeias, i.e. “words that originate as imitations of sounds” (Dingemanse 2019: 28). This does not necessarily imply that all UGC onomatopoeias also are ideophones.

<sup>5</sup> For verbs such as CC *ñemé* ‘chew’ (5b) or Santiago Capeverdean *fasi* ‘make’ (6a) and *odja* ‘see’ (6b), which have transitive uses, one could also argue that the ideophones fill an object slot and are not adverbs. However, such ideophones do not behave as prototypical nouns, as (i) they cannot combine



## (6) Santiago Capeverdean

(a) *fasi* ‘make’ + *menhi-menhi* (IDEO) > *fasi menhi-menhi* ‘have **interference**, be **jammed** (television)’

(b) *odja* ‘see’ + *blu* (IDEO) > *odja blu* ‘see **pink elephants**, have **alcohol-related hallucinations**’

UGC ideophones match the typological characterization elaborated by Dingemanse in 2012, who defines ideophones as “marked words that depict sensory imagery” (Dingemanse 2012: 655). He refined the definition of ideophone in 2019, as a “member of an open lexical class of marked words that depict sensory imagery” (Dingemanse 2019: 16). While it is indisputable that ideophones form an open lexical class in many Niger-Congo languages (e.g. in Koalib (kib), see Quint 2018), this property is not necessarily shared by all UGCs, in particular insular UGCs, hence our preference for the 2012 definition.

### 3.2. Previous studies on UGC ideophones

UGC ideophones have already been studied in some detail for Guinea-Bissau Creole<sup>6</sup> and briefly mentioned for Capeverdean (Quint 2000b: 107-109; 2008: 72-75) and Casamance Creole (Biagui 2018: 265, 372-373; Biagui & Quint 2013: 47; Dalphinis 1986: 107-108). This material mainly consists of lexeme lists with brief comments. Furthermore, only three of the cited authors have published lists of more than 20 ideophones: Kihm (1995) provides 35, Scantamburlo (1999) 51 and Biagui (2018) 62 ideophones. This paper aims at providing a more complete coverage of the category of ideophones across Upper Guinea Creoles.<sup>7</sup>

---

with nominal determiners (such as demonstratives, numerals, adjectives, etc.) and (ii) generally cannot occupy a nominal slot in a sentence (although the specific behavior of *menhi-menhi* is more ambiguous to this respect, criterion (i) is still valid for this item). Hence, in a sequence such as *odja blu* (6b), from a syntactic point of view, *blu* behaves like a non-ideophonic adverb such as ‘well’ or ‘clearly’ and not like a noun such as ‘car’, ‘water’ or ‘elephant’.

<sup>6</sup> See Quint (2023: 460-461), Truppi & Costa (2019), Scantamburlo (1981: 66-67; 1999: 189-191), Couto (1994: 102-104; 1995: 207-215), Childs (1994: 265-266, 277), Kihm (1994: 76-78), Doneux & Rougé (1988: 27-28), Mbodj (1979: 79; 1984: 58) and Wilson (1962: 62).

<sup>7</sup> Papiamentu (i.e. the third branch of UGCs, see §2.1) is not taken into account. Indeed, as far as we can tell, Papiamentu does not seem to have an ideophonic word class comparable to what will be described hereafter.

#### 4. Data provenance and methodology

This study builds on two main datasets, for continental and insular UGCs respectively.

##### 4.1. The continental dataset

For the continental Upper Guinea Creoles dataset, the following varieties are taken into account (Map 1): Bissau, Cacheu and Geba (Guinea-Bissau) and Casamance (Senegal). In practice, the Casamance variety was used as the standard against which other varieties are compared. As Biagui, himself a Casamance Creole native speaker, had already collected 81 ideophonic items in his dictionary (Biagui in preparation) – the longest list so far compiled for any UGC variety – Quint used these items and systematically checked their semantic equivalents for all other continental varieties. For this, several native speakers and available published sources, mainly Scantamburlo<sup>8</sup> (1999), were consulted (Table 1). All CC ideophones and their other continental UGC counterparts are given in Appendix 1 (Tables 1.1 to 1.5).

Variety	Source	Number of consultants <sup>9</sup>	Code
Casamance	Biagui (to appear) + personal knowledge	1	CC
Geba	fieldwork (2019)	1	Geba
Cacheu	fieldwork (2019)	3	Cacheu1/2/3
Bissau	fieldwork (2019)	2	Bissau1/2
	Scantamburlo (1999)	1	Bissau3

**Table 1:** Origin of the data for each UGC variety.

The main reasons for the choice of Casamance Creole as the point of departure for the study of these four continental varieties were (i) the availability of a dataset compiled earlier by Biagui and (ii) the fact that he could himself re-check this data and provide in-depth insights as to the behavior of continental UGC ideophones.

<sup>8</sup> In two cases at least (\*far/ and \*/caŋ/, see Tables 1.3 and 1.4 in Appendix 1), we also resorted to Scantamburlo (2002) to check ideophones that were not listed in Scantamburlo (1999).

<sup>9</sup> Regarding Scantamburlo’s dictionary, the data obviously come from more than one consultant. However, as it is a single written source and the specific contribution of each consultant is not mentioned, we will make the approximation of considering that this publication represents one dataset, and therefore is the equivalent of one consultant.

The main reason for the choice of the remaining varieties (i.e. Bissau, Cacheu and Geba) was the desire to provide an exhaustive coverage of continental UGCs, as these varieties, together with Casamance Creole, are the only continental varieties dating back to the settlement period that are still spoken today (see also Quint & Moreira Tavares 2019: 118).

#### **4.2. *The insular (Capeverdean) dataset***

For Capeverdean, we checked two varieties, namely Santiago and Fogo (Map 2), and looked for ideophonic lexemes comparable to the ones found by Biagui in Casamance Creole. For Santiago, we mostly drew on Quint's own lexical databases, whether published (Quint 1998; 1999) or not. A secondary source was Rougé (2004). The Santiago ideophones considered in this study are given in Appendix 2 (Table 2.1). For Fogo, we relied on the material gathered by Moreira Tavares that was published in Quint & Moreira Tavares (2019). Furthermore, for the sake of comparability, the semantic equivalents of all 81 ideophones attested in Casamance Creole were also systematically checked with one Santiago native speaker.

The main reasons for the choice of Santiago and Fogo varieties were (i) availability of data and (ii) the fact that both Santiago and Fogo varieties are Sotavento (southern) Capeverdean, in which the African element is particularly important, when compared with northern Capeverdean (Quint 2000b: 71-97). Assuming that UGC ideophones were probably an African feature (an intuition that was confirmed by the present study, as shown below), it appeared preferable to choose more Africanized Capeverdean varieties.

#### **4.3. *The overrepresentation of verb intensifiers***

In this study, we have clearly favored a semantic subcategory of ideophones, namely verb intensifiers (see also §3.1). They represent 61 items out of a list of 81 Casamance Creole ideophones prepared by Biagui, i.e. 75% of the total. The main reason underlying this imbalance is the specific semantic relation between a verb-intensifying ideophone and the verb it modifies. A given verb-intensifying ideophone rarely collocates with more than one or two verbs and, conversely, a given verb is rarely intensified by more than one or two different ideophones. Consequently, the semantic relation between a verb and its intensifying ideophone is salient and clearly

perceived by most native speakers of any UGC variety, which also renders verb-intensifying ideophones easier to elicit for field linguists, when compared with other semantic types of ideophones. In our database, we see, for instance, that several sound ideophones can be associated with the action of FALLING, according to the type of noise produced by this action, e.g. *furba* (in rice bran, see (5a)), *bup* (someone on a hard surface), *tim* (a metallic object). In other words, verb-intensifying ideophones can be considered, at least in UGCs but also in other languages (e.g. in Koalib, see Quint 2018), as a prototypical semantic type of ideophones, which probably accounts for the fact that our comparative database has a strong bias towards verb-intensifying ideophones. As a matter of fact, all previous work devoted to UGC ideophones also mentions a majority of verb-intensifying ideophones in the examples and lists.

Furthermore, when we began this comparative work, we were mostly focused on intensifying ideophones and tended to discard other semantic types from the ideophone category, in particular sound ideophones. For instance, the equivalents of Casamance Creole sound ideophones *furba* (5a) and *cákum-cákum* (5b), which Biagui (forthcoming) had already recorded in his dictionary of Casamance Creole, were not systematically checked for other varieties and could therefore not be taken into consideration in this comparative study.

## 5. Ideophones in the continental branch

### 5.1. Phonology

As observed in many Niger-Congo languages (Creissels 1994: 30, Quint 2006: 101; 2009a: 91; 2018: 181-190), the phonological characteristics of continental UGC ideophones significantly differ from other lexical items of these languages. This difference is conspicuous at least at four levels: prosody, syllabic structure, vowel pattern and consonant inventory.<sup>10</sup>

#### 5.1.1 Prosody

Nearly all continental UGC verb-intensifying ideophones are pronounced with a clearly recognizable high tone which extends across all syllables of the word. This

---

<sup>10</sup> These four levels are explicitly mentioned by Dingemanse (2012: 656) as typically marked for many-ideophones across the languages of the world.

(H)<sub>n</sub> all-high tone profile is most plausibly a Mandinka feature (§5.4). It contrasts with all other lexical items of the language, which follow a stress pattern (inherited from Portuguese). Consider again the verb + ideophone sequence given in (1), here renumbered as (7).

- (7) Bissau1 *braŋku* ‘be white’ + *fandaŋ* (IDEO) > *braŋku fandaŋ* ‘be **very** white, be **as white as snow**’

The verb *braŋku* [ˈbraŋku] ‘be white’ (< Port. *branco* [ˈbrẽku]), which is a non-ideophonic lexical item, displays an S-U (=stressed–unstressed) stress pattern. The ideophone *fandaŋ* [fándáŋ]<sup>11</sup> ‘very’ (non-Romance), which is an ideophonic lexical item, displays a H-H (high-high) tone pattern.

However, some continental ideophones do have a stress (*not* tonal) prosodic pattern. In (8), in the sequence *ñemé cákum-cákum* [ɲeˈme ˈcakum ˈcakum] both the verb *ñemé* ‘chew’ and the sound ideophone *cákum-cákum* are stressed.

- (8) CC *ñemé* ‘chew’ + *cákum-cákum* (IDEO) > *ñemé cákum-cákum* ‘chew **noisily**’

In continental UGCs, most stressed ideophones seem to belong to semantic categories other than verb intensifiers.<sup>12</sup> Considering the overall prosodic system of the language, a stressed ideophone is less marked than a tonal ideophone, which confirms the prototypical character of verb intensifiers by contrast with other types of ideophones.

### 5.1.2 Syllabic structure

As shown in Table 2, all continental UGC ideophones save one end in a consonant and over two thirds (69%) are monosyllabic.<sup>13</sup> This data is obviously at variance with the well-formedness standard for continental UGC words, where (i) disyllabic items are clearly favoured, as shown by the fact that 65% of Casamance Creole lexical words

<sup>11</sup> In the phonetic transcriptions provided in this paper, an acute accent above a vowel, e.g. [V́], indicates a high tone.

<sup>12</sup> However, there is at least one case of a stressed verb intensifier ideophone in continental creoles: see fn. 32 and 40.

<sup>13</sup> The peculiar syllabic profile of continental UGC ideophones was already noticed by other authors, e.g. Couto (1995: 212), Kihm (1994: 76) and Wilson (1962: 34).

(and 39% of all lexical items<sup>14</sup> of the language) have a C(C)V.C(C)V structure (Biagui 2018: 89), and (ii) C(C)V is the dominant syllable type, with 73% in Casamance Creole (Biagui 2018: 86) and 65,79% in Bissau Creole (Mbodj 1979: 54).

Syllabic structure	Number of ideophones	%	Syllables	Number of ideophones	%
CV	1	1%	1	56	69%
CVC	55	68%			
CV.CVC	15	19%	2	21	26%
CVC.CVC	6	8%			
CV.CV.CVC	2	2%	3	4	5%
CVC.CV.CVC	2	2%			
<b>Total</b>	<b>81</b>	<b>100%</b>		<b>81</b>	<b>100%</b>

**Table 2:** Syllabic structure of continental UGC ideophones (based on Biagui’s data for Casamance Creole).<sup>15</sup>

### 5.1.3 Vowel patterns

Polysyllabic continental UGC ideophones display a special type of vowel harmony: all vowels of a given ideophone are the same, as in (9) and (10). The only exception is seen in (12).

(9) CC *pagá* ‘switch off’ + *kamaj* (IDEO) > *pagá* [o] *kamaj* ‘switch [sth.] off **suddenly**’

(10) Cacheu1 *beju* ‘be old’ + *kokorot* (IDEO) > *beju kokorot* ‘be **very** old, be as old as **Methuselah**’

### 5.1.4 Consonant inventory

As a rule, continental Upper Guinea Creoles have no voiced fricatives in their core vocabulary<sup>16</sup> (Quint 2023; Biagui 2018: 55, 79; Kihm 1994: 17-18; Doneux & Rougé

<sup>14</sup> Throughout this paper, we make a distinction between “lexical word”, a category including primarily open lexical classes (nouns, verbs...), “morphological (or grammatical) word”, a category including closed lexical classes (adpositions, conjunctions...) and “lexical item”, a general label including any word of the language, whether “lexical” or “morphological”.

<sup>15</sup> The syllabic structure of the other continental UGC varieties is not significantly different.

<sup>16</sup> This trend can also be observed in conservative varieties of Santiago Capeverdean (an insular UGC), see Quint (2000a: 112-114) and §4.1.4.

1988: 6). However, /v/ and /z/ are attested for several varieties in our ideophone data, as shown in Table 3.

Ideophone						
No.	CC equivalent	Voiced fricative	Variety and consultant	No. of occurrences	Semantic type	Associated verb(s)
(1)	<i>bap</i>	<i>vap</i>	Bissau1	1	sound	<i>sintá</i> ‘sit’ + <i>kay</i> ‘fall’
(2)	<i>kamaj</i>	<i>vut</i>	Bissau1	1	immediate	<i>pagá</i> ‘switch off’
(3)	<i>tar</i>	<i>vap</i>	Cacheu3	1	sound/intensifier	<i>bafatiyá</i> ‘slap’
(4)	<i>bup</i>	<i>vip</i>	Cacheu1 + Cacheu3	2	sound	<i>kay</i> ‘fall’
		<i>vup</i>	Geba	1		
(5)	<i>lip</i>	<i>vip</i>	Geba	1	intensifier	<i>pisadu</i> ‘be heavy’
(6)	NA <sup>17</sup>	<i>zip</i>	Geba	1	intensifier	<i>gros</i> ‘be fat/thick’
			<b>TOTAL</b>	<b>8</b>		

**Table 3:** Occurrences of continental UGC ideophones containing a voiced fricative.

All in all, /v/ and /z/ appear very rarely in our data (8 occurrences for ca. 500 UGC ideophones, i.e. 1%) and in variants that have very similar phonological shapes, i.e. /vVp/ ~ /vVt/ ~ /zVp/. However, voiced initial fricatives are attested in all continental UGC varieties except Casamance Creole, which, admittedly, is the most basilectal of the group (see Biagui et al. forthcoming; Quint 2023: 451-452, 454, 457). Note also that at least three of the six ideophones for which voiced fricatives are attested are (or can be used as) sound ideophones, which represent only a minority of our sample (§4.3): the sporadic use of voiced initial fricative variants in sound ideophones may be linked with ICONICITY (Dingemanse 2012: 657-660), in this case the necessity to depict as faithfully as possible the sound perceived by the speakers.<sup>18</sup>

### 5.1.5 Lexical reduplication

Contrary to many other languages (Dingemanse 2012: 656; Quint 2018), in continental UGCs, only a small number of ideophones have a reduplicated base form. In our

<sup>17</sup> Some few ideophones were taken into consideration even though they were missing in our CC list, for more details see §5.3.

<sup>18</sup> Other languages have been reported to show a higher frequency of voiced fricatives in ideophonic items than in the overall lexicon. This is for example the case of /z/ ~ /dz/ (both realizations appearing in free variation) in Pastaza Quichua (qvz) (Nuckolls et al. 2016).

Casamance Creole sample of 25 polysyllabic ideophones (21 disyllabic + 4 trisyllabic, see Table 2), only six ideophones (i.e. approximately a quarter of the total number of polysyllabic ideophones and  $6/81 = 7\%$  of all continental ideophones) have a fully (one item, see (11)) or partially (five items, see (10, 12)) reduplicated lexical base form.

(11) CC *termé* ‘shake/shiver’ + *bok-bok* (IDEO) > *termé bok-bok* ‘shake like a leaf’

(12) Geba *jantí* ‘hurry up’ + *fit-fat* (IDEO) > *jantí fit-fat* ‘hurry up making great strides’

## 5.2. Morphosyntax<sup>19</sup>

### 5.2.1 Word order

In all continental UGCs, the ideophone always follows the word (generally a verb) whose meaning it modifies (see also Kihm 1994: 76). To the best of our knowledge, no other element can be inserted between an intransitive verb and an ideophone (13)-(14).

(13) CC

*gatu kay furbap déntur di bidon*

cat fall.PFV IDEO inside of barrel

‘The cat fell **with a soft noise** inside the barrel [full of rice bran].’

(14) CC

*kontrá ku sol kinti Mariya simí fes*

when SUB sun be.hot.PFV Mary disappear.PFV IDEO

*ninjeŋ ka torná wojá-l más*

nobody NEG do.again.PFV see.PFV-O3SG anymore

‘During the hottest part of the day, Mary disappeared **completely** [and] nobody saw her anymore.’

With transitive verbs, the object is usually inserted between the verb and the ideophone. This object can be either a noun (15) or a pronoun (16)-(17).

---

<sup>19</sup> In this part, most examples are drawn from Casamance Creole, which Biagui speaks natively (§4.1). Our comparison of continental UGC varieties was mainly limited to the lexicon and, in most cases, the available data does not allow us to say whether the morphosyntactic processes we describe hereafter are limited to CC or extend across all continental UGCs.



## (15) Bissau

*iransegu kumé karnel buk*  
python eat.PFV sheep IDEO

‘The python swallowed the sheep **whole**.’ (Kihm 1994: 76)<sup>20</sup>

## (16) CC

*i kumé-l bik*  
s3SG eat.PFV-O3SG IDEO

‘S/he ate it **all**.’

## (17) CC

*disna k-e jusiyá e ta pasá ηútur maj-maj*  
since SUB-S3PL quarrel.PFV S3PL HAB go.by each.other IDEO.RED<sup>21</sup>

‘Since they quarrelled, they don’t talk anymore (lit. since they quarrelled, they go by each other **with a total lack of attention**).’

## 5.2.2 Modified elements

In an overwhelming majority of cases, the ideophone modifies a verb (see all examples above) or a qualifier (i.e. an adjective behaving as a verb when used in predicative function,<sup>22</sup> see Quint 2023: 457, 459). However, some other parts of speech, such as adverbs (18) and pronouns (19), may also be modified by an ideophone.

(18) CC *didiya* ‘at noon’ + *lɛm* (IDEO) > *didiya lɛm* ‘at **solar noon**, when the sun is just **at its peak/zenith**’

(19) Cacheu1 *a-mi* ‘me’ + *kondonj* (IDEO) > *a-mi kondonj* ‘**on my own**, being (myself) **hopelessly alone**’

In (19), *kondonj* could be itself analysed as a non-ideophonic adverb meaning ‘alone’, but this same element also combines with verbs (20) in prototypical ideophonic

<sup>20</sup> This usage of *buk* in association with the verb *kumé* ‘eat’ is not listed in our own database.

<sup>21</sup> For the use of expressive reduplication in Casamance Creole, see §5.2.3.1.

<sup>22</sup> Note that qualifiers can be modified by an ideophone both in predicative and attributive function.

constructions where one can see that the notion of LONELINESS that *kondoŋ* conveys in (19) is linked with the notion of EMPTINESS or LACK OF CONTENT:

(20) Bissau

*garafa*     *linpu*                     *kondoŋ*  
bottle     be.clean/empty.PFV   IDEO  
'The bottle is **totally** empty.' (Kihm 1994: 77)

The existence of (20) justifies the analysis of *kondoŋ* in (19) as an ideophone modifying the pronoun.

### 5.2.3 Morphology

Although UGC ideophones are clearly adverbs, they can be involved in at least three types of morphological processes: expressive reduplication, vowel lengthening and ideophone-to-verb derivation.

#### 5.2.3.1 Expressive reduplication

Expressive reduplication is used to strengthen the meaning of a given ideophone (17, 21, 22).

(21) CC *burmeju* 'red' + *wək* (IDEO)

- > *burmeju wək* 'vivid/bright red'
- > *burmeju wək-wək* 'extremely bright red'

(22) CC *susu* 'dirty' + *potok* (IDEO)

- > *susu potok* 'very dirty'
- > *susu potok-potok* 'horribly dirty'

Ideophones whose lexical form is reduplicated (§5.1.5) can also undergo (partial) expressive reduplication (23).

(23) CC *termé* 'shake/shiver' + *bok-bok* (IDEO)

- > *termé bok-bok* 'shake like a leaf'
- > *termé bok-bok-bok* 'shake like a leaf without stopping'

## 5.2.3.2 Final vowel lengthening

Vowel lengthening is another means to increase the expressiveness of an ideophone and it is only attested for trisyllabic ideophones. In all known cases, it is the last vowel that is lengthened (24):

- (24) CC *saŋ* ‘healthy’ + *keŋkereŋ* (IDEO)  
 > *saŋ keŋkereŋ* ‘very healthy’  
 > *saŋ keŋkereeŋ* ‘extremely healthy’

## 5.2.3.3 Ideophone-to-verb derivation

Several Casamance ideophones can be derived into verbs by the addition of the stressed suffix *-/i/* (25)-(26).

(25) CC

- (a) *pañá* ‘thicken’ + *təkɛp* (IDEO)  
 > *pañá təkɛp* ‘get **very** thick’ (glue, sauce)  
 > *təkɛpí* ‘get tough’

- (b) *e añju təkɛpí suma turu nobu*  
 DEM.PROX new.born.baby **get.tough**.PFV like bull young  
 ‘This newborn baby has a body as **tough** as a young bull.’

(26) CC

- (a) *termé* ‘shake/shiver’ + *bok-bok* (IDEO)  
 > *termé bok-bok* ‘shake **like a leaf**’  
 > *bok-bokí* ‘shake like a leaf’

- (b) *bu na bok-bokí na suma miñjer bej-a*  
 S2SG IPFV.PROG **shake** ASS like woman old-F  
 ‘You are **shivering** like an old woman.’

The suffix *-/i/* is very productive in the language and can also be used to derive verbs from other parts of speech, e.g. nouns (27), including borrowings:

- (27) CC *fotó* [fo'to] ‘photograph’ (< French *photo*)  
 > *fotoí* [foto'i] ‘take a photo/picture [of s.o./sth.]’

Therefore, the ideophone-to-verb derivation fits into the general morphological framework of Casamance Creole.

### 5.3. Diatopic variation

#### 5.3.1 Shared ideophonic roots

If we look at the number of shared ideophonic roots between the different UGC varieties (Table 4), taking into account that Casamance Creole was considered as the standard for comparison, we can distinguish four types of distribution profiles, which are discussed hereafter.<sup>23</sup>

Distribution profile	No. of items	Distribution type	No. of items	%
CC-only	22	CC-only	22	26%
CC + 1 <sup>24</sup>	4	Minority models	14	17%
CC + 2	5			
CC + 3	5			
CC + 4	11	Common core	39	46%
CC + 5	8			
CC + 6	17			
CC + 7	3			
CC vs. Guinea-Bissau	3	Other distribution	9	11%
no CC vs. Guinea-Bissau	3			
(CC + Guinea-Bissau) vs. Guinea-Bissau	3			
<b>Total</b>	<b>84</b>		<b>84</b>	<b>100%</b>

**Table 4:** Number of UGC continental varieties sharing the same ideophonic root as Casamance Creole (CC).

In the ‘*CC-only*’ type, 22 CC ideophones (i.e. 26% of the total) have either no semantic counterpart in any other UGC variety or their respective UGC semantic equivalents are not attested in more than one UGC variety each and these UGC semantic

<sup>23</sup> This classification into distribution types is also used in the data appendices.

<sup>24</sup> The digit following ‘CC + ’ indicates the number of consultants from other UGC continental varieties sharing an ideophonic root with Casamance Creole (CC), e.g. ‘CC + 1’ means that one consultant speaking a continental UGC variety other than CC provided an ideophone whose phonological form is comparable to the CC semantically equivalent ideophone.

equivalents have a lexical root different from CC. The ‘CC-only’ distribution type underlines the outlier status of Casamance Creole among continental UGCs (Quint & Moreira Tavares 2019: 133-134), probably due to (a) a peculiar Jola and Nyun<sup>25</sup> adstrate and (b) the political boundary, which, for almost 140 years (Biagui & Quint 2013: 41, Biagui 2018: 18; Quint 2023: 451-452, Biagui et al. forthcoming), has been isolating Casamance Creole – spoken in Senegal – from the remaining continental UGCs – all spoken in Guinea-Bissau.

The contrast between CC and the other UGCs is also conspicuous at the phonological level, even when the CC form is obviously related to other UGCs, i.e. for ideophones belonging to the ‘common core’ type, see entries (30) and (37) in Table 1.3 in the appendix.

(28) Continental UGCs *burmeju* ‘red’

> Geba + Cacheu1/2/3 + Bissau1/3 *burmeju wak* ‘**bright** red’

≠ CC *burmeju wək*

(29) Continental UGCs *moli* ‘soft’

> Geba *moli pɔcɔk* ~ Cacheu1/2/3 + Bissau1 *moli pɔtɔk* ~ Bissau3 *moli potok*  
‘**very** soft’

≠ CC *moli botok*

In the ‘*common core*’ type, nearly one half (39/81 = 48%) of our standard list of 81 CC ideophones have a close equivalent for at least 4 of the 7 consultants (§4.1) from other UGC varieties. We can therefore reasonably posit that these forms can be traced back at least to proto-continental UGC, i.e. to the second half of the 16<sup>th</sup> century (§2.2).

In the ‘*minority model*’ type, some CC-related ideophones were found in the data of only a minority (i.e. fewer than four) of consultants from other UGC varieties. In many cases, these common roots are probably also traceable to proto-continental UGC, as it is very plausible that such phonological and semantic coincidences between several UGCs are not due to mere chance. However, a lower number of formal matches implies a higher degree of uncertainty. It is also striking that the ‘*minority model*’ type concerns fewer items than the ‘*common core*’ and ‘*CC-only*’ types, which means that the main split is between ideophones only found in CC and those shared by all

---

<sup>25</sup> Jola and Nyun are not provided with an ISO Code-639, because there are no satisfactory code for each of these language groups.

(or most) UGCs. The ‘minority model’ therefore represents an intermediary category, whose existence may be attributable to factors related to the fieldwork setting (e.g. misunderstandings between collector and consultants) or the lower frequency of forms.

The ‘*other distribution*’ type covers three subtypes which deserve to be studied in more detail. In the ‘CC vs. Guinea-Bissau’ subtype, the Casamance Creole (CC) form contrasts with a form attested in several Guinea-Bissau UGC varieties (30).

(30) Continental UGCs *tesu* ‘be solid’

- > Geba + Cacheu2 + Bissau1 *tesu **kaŋ*** ‘be **very** solid’ ~ Cacheu3 *tesu **taŋ***
- ≠ CC *tesu **liŋ***

In the ‘no CC vs. Guinea-Bissau’ subtype, a form attested in several Guinea-Bissau UGC varieties is lacking in Casamance Creole (31).

(31) Continental UGCs *negá* ‘refuse’

- > Geba + Cacheu1/2 + Bissau3 *negá **far*** ‘refuse **adamantly**’
- ≠ no known form in CC<sup>26</sup>

Finally, in the ‘(CC + Guinea-Bissau) vs. Guinea-Bissau’ subtype, the CC form clusters with one or several Guinea-Bissau forms and contrasts with the remaining Guinea-Bissau forms (32).

(32) Continental UGCs *sukuru* ‘be dark’

- > CC + Geba *sukuru **mut*** ‘be **pitch/completely** dark’
- ≠ Cacheu3 + Bissau2/3 *sukuru **mip***

### 5.3.2 Variation according to variety

Except for the distinctive character of CC reflected in the ‘CC-only’ type, which is also linked with the choice of CC as a standard of comparison, there are no significant differences between the other UGC varieties as regards their ideophone inventories.

Bissau2 and Bissau3 are the only two datasets lacking an equivalent for a majority of the 81 CC ideophones (Table 5). However, these results are an experimental

---

<sup>26</sup> This subtype explains why the total number of ideophones in Table 4 is 84 instead of 81 elsewhere. Here, 84 results from the 81 ideophones listed for CC (§4.1) plus the 3 ‘no CC vs. Guinea-Bissau’ cases.

artifact, because we did not check the complete list of CC ideophones with consultant Bissau2. As a matter of fact, with consultant Bissau2, we only recorded some forms we wanted to check after having worked with Bissau1. The Bissau3 dataset was taken from a publication (Scantamburlo 1999: 189-191, see §4.1), and we could not check missing forms with a consultant.

If we consider the other consultants and varieties, their behavior is quite similar: they lack equivalents for roughly the same proportion of the total list (between 23 to 36 items, i.e. 28% to 44% of the total) and most of them have few or no semantic equivalents for the ideophones belonging to the ‘CC-only’ type (§5.3.1). Consultant Cacheu3 has the lowest number of missing CC equivalents, which means Cacheu3 has more in common with CC than any other consultant. The specific proximity between CC and Cacheu3 is reflected by several examples, such as (33).

(33) *pisadu* ‘be heavy’

> *pisadu sip* (Geba + Cacheu1/2/3 + Bissau1/2) ‘be as heavy as lead’

≠ *pisadu lip*<sup>27</sup> (CC + Cacheu3).

One could explain this higher degree of similarity by the fact that the Cacheu and Casamance UGC varieties are historically related (§2.1). Nonetheless, the other available datasets for Cacheu (Cacheu1 and Cacheu2) do not display noticeable differences to Geba and Bissau (or rather Bissau1, the only complete dataset available for Bissau) regarding their number of missing equivalents of CC ideophones. Hence, it is not possible to contrast the ideophone inventories of Cacheu with that of Geba and/or Bissau. UGC ideophones are definitely very similar in all Guinea-Bissau varieties, which is due to their common origin (see §2.2 and §5.3.1) but also to the levelling influence of the Bissau variety, which is about to erase the Cacheu and Geba varieties (§2.1).

---

<sup>27</sup> Note that *lip* is only attested as an intensifier of *fartá* ‘be satisfied’ in Cacheu3 and that Cacheu3 also uses *sip* (as other Guinea-Bissau continental varieties). However, given that (i) *lip* is not used by any other Guinea-Bissau variety; (ii) SATISFIED and HEAVY are obviously semantically related; and (iii) the local Cacheu variety is about to be totally replaced by the Bissau variety (§2.1), this specific resemblance between CC and Cacheu3 can be considered as linguistically significant.

Missing equivalents <sup>28</sup>		
Variety	Total	CC-only
Geba	30	22
Cacheu1	29	21
Cacheu2	36	18
Cacheu3	23	15
Bissau1	31	18
Bissau2	74	22
Bissau3	47	22

**Table 5:** Number of missing equivalents of CC ideophones in other UGC continental varieties.

#### 5.4. Origin

Most of the ideophones whose origin is identified come from Mandinka and/or Wolof (see Table 6 and Table 3.1. in Appendix 3).

Continental UGC ideophones		Ideophones with an attested African etymon			
Distribution type	Total	Wolof	Mandinka	Wolof or Mandinka	Total
CC-only	22	4	2	1	7
Minority models	14	5	3	0	8
Common core	39	2	10	5	17
Other distribution	9	4	1	2	7
<b>Total</b>	<b>84</b>	<b>15</b>	<b>17</b>	<b>7</b>	<b>39</b>

**Table 6:** African-derived continental UGC ideophones according to their distribution type as defined in Table 4.

A significant proportion ( $39/84 = 46\%$ ) of the ideophones studied in this paper have a plausible African etymon. This confirms Couto’s (1995: 212) and Kihm’s (1994: 76)

<sup>28</sup> We forgot to check the equivalents of some few CC ideophones in Cacheu2 (four items) and Bissau1 (three items), which may account for the fact that these two datasets have a total number of missing equivalents slightly higher than Geba, Cacheu1 and Cacheu3.



intuition that continental UGC ideophones originally come from West African languages.<sup>29</sup>

Wolof (Niger-Congo, Atlantic) and Mandinka (Niger-Congo, Mande) are the only known African sources for continental UGC ideophones. This result is in accordance with the fact that these languages have long been recognized as the main African substrates of UGCs (§2.2). Nevertheless, the fact that Mandinka and Wolof, due to their widespread use as vehicular languages in West Africa, are better described than many other languages of the region – see, in particular, the detailed lexicographic works for Mandinka by Creissels (2012) and Creissels et al. (1982), and for Wolof by Diouf (2003) and Fal et al. (1990) – is also a significant factor accounting for the absolute dominance of these two languages as etymological sources of UGC ideophones.

The ratios of Mandinka- and Wolof-derived continental UGC ideophones are roughly equivalent. This result is at variance with what was found in previous studies that considered UGC African-derived items in all parts of speech (Quint 2000b: 23-34, 110-113; Quint 2008: 32-47; Quint & Moreira Tavares 2019: 124-126). These studies showed a clear dominance of Mandinka-derived items over Wolof-derived items with at least a 2:1 ratio. In other words, continental UGC ideophones seem to be more Wolofized than the other African-derived items of the language. However, if we look at Table 6 in more detail and check the ratio of Mandinka and Wolof for each of the four recognized distribution types, a more nuanced picture emerges.

Mandinka is clearly dominant in the ‘common core’ type (10 Mandinka-derived vs. 2 Wolof-derived ideophones), i.e. among the ideophones which can be traced back to proto-continental UGC with most certainty. Conversely, Wolof seems to dominate in all other types.

The ‘CC-only’ type comprises the ideophones considered as typical of Casamance Creole (CC). As CC is the northernmost variety of continental UGC, as the Wolof-speaking area is situated north of the Mandinka speaking area, and as CC is spoken in Senegal where Wolof is the main lingua franca, it comes as no surprise that the Wolof element is stronger in CC than in any other continental UGC.<sup>30</sup>

---

<sup>29</sup> Conversely, it goes against Childs (1994: 265-266), who, after a first exploratory study on Guinea-Bissau Creole, wrote that “[n]one of the [Guinea-Bissau Creole] ideophones were identified in the first languages [=Atlantic and Mande traditional languages] used alongside Guinea-Bissau Creole” and inferred from this observation (which is contradicted by the present data) “that the ideophones were generated internally, rather than borrowed”. Wilson (1962: 34) basically upholds Child’s position.

<sup>30</sup> For the question of Wolof influence on CC (and more generally on northern continental UGCs), see Biagui et al. (forthcoming) and Quint & Moreira Tavares (2019: 124-126, 135-139).

The ‘minority models’ assemble ideophones that are less likely to be traced back to proto-continental UGC than those of the ‘common core’. The ‘other distribution’ type is characterized by the fact that not all UGCs share the same lexical root for a given ideophone, which implies that, for most of these cases, at least one of these roots cannot be traced back to proto-continental UGC.

In sum, the Mandinka component is stronger among those ideophones that can be traced back to proto-continental UGC, i.e. the common stage underlying all known continental UGCs. This seems to suggest that the particularly high ratio of Wolof-derived items among continental UGC ideophones is due to the fact that a significant part of these Wolof-derived items made their way into some UGC varieties *after* the period of formation of continental UGCs, i.e. Wolof is both a *substrate* for UGCs – it contributed some ideophonic elements during their period of formation – and an *adstrate*, as it contributed some ideophonic elements after the period of formation. The same applies to Mandinka, which is both a substrate for all UGCs and an adstrate for several contemporary continental UGCs (Quint & Moreira Tavares 2019: 125-126). However, the results given in Table 6 suggest that Mandinka has been a less powerful adstrate than Wolof.<sup>31</sup>

Another argument in favor of the prevalence of Mandinka as a continental UGC substrate is the fact that in Mandinka – a tonal language – ideophones are generally characterized by the same (H)<sub>n</sub> all-high tone profile (Creissels 2013: 323; 2012; Creissels et al. 1982: xvii) as continental UGC ideophones (§5.1.1). As Wolof is a stress language, it is probable that the tonal profile of UGC continental ideophones was inherited from Mandinka.<sup>32</sup>

Besides Mandinka and Wolof, at least one continental UGC ideophone is derived from Portuguese (34).

(34) Continental UGCs *kurpu limpu* ‘naked’, lit. ‘body clean’

> Geba + Cacheu1/3 + Bissau1 *kurpu limpu nu* ‘stark naked’

= CC *kurpu limpu nuŋ*

with UGC *nu/nuŋ* < Port. *nu* [ˈnu] ‘naked’

---

<sup>31</sup> The relative weakness of Mandinka as an adstrate may also be partly explained by the fact that CC (the most Wolof-influenced continental UGC variety, see Quint (2023: 451-452), Biagui et al. (forthcoming)) was chosen as a standard for this comparative study.

<sup>32</sup> Most notably, [ˈmoku], the only known UGC verb-intensifying ideophone having a clear stress pattern on the continent (see fn. 40 and Table 3.1 in Appendix 3), comes from Wolof.

## 6. Ideophones in the insular branch (Capeverdean)

In this section, a sample of 34 insular UGC ideophones will be systematically compared with their continental counterparts.

### 6.1. Phonology

#### 6.1.1 Prosody

Contrary to continental UGCs, insular UGC ideophones are prosodically much more integrated into the general rules applying for the language. All insular ideophones are stressed (like any other lexical word in all insular UGCs), and two stress profiles are allowed, oxytonic (35) and paroxytonic (36).

(35) Santiago *abri* ‘open’ + *uandaj* [wẽ'dẽ] (IDEO) > *abri uandaj* ‘open **wide** [a door]’

(36) Santiago *sai* ‘get out’ + *futi* [ˈfuti] (IDEO) > *sai futi* ‘get out **at once/quickly**’

The paroxytonic profile is clearly dominant, as it represents 24.5<sup>33</sup> items (i.e. 82%) out of a total of 30 polysyllabic ideophones.

#### 6.1.2 Syllabic structure

Contrary to continental UGCs, a majority of insular UGC ideophones (72%) contain only C(C)V open syllables (Table 7), and most of them (63%) are disyllabic (Table 8). Insular UGC ideophones therefore exactly fit into the syllabic canon of the language, which is characterized, at least in the two varieties under study, by a preponderance of C(C)V syllables and a majority of disyllables for stressed items (see Quint 2000a: 35-40 for Santiago and Moreira Tavares 2020: 97-101 for Fogo).

Another peculiarity of insular UGC ideophones is the fact that at least six of them (i.e. roughly one out of six) contain one or two CCV syllables, while these complex onsets are not attested in continental UGC ideophones. Actually, CCV syllables are

---

<sup>33</sup> The reason why, in this section devoted to the phonology of insular UGC ideophones, we are dealing with half-words is due to the fact that one of these ideophones has at least three different variants (*txulupu* [tʃuˈlupu] ~ *txulufu* [tʃuˈlufu] ~ *txufun* [tʃuˈfũ]) with two different stress and syllable patterns (/CVˈCVCV/ (= *txulupu* + *txulufu*) vs. /CVˈCVC/ (= *txufun*)), see entry (31) in Table 2.1 in the appendix. In consequence, we have considered that each of these two patterns counted as 0.5 item in our word counts.

much more frequent in Capeverdean (12% in texts, see Quint 2000a: 393, 396, 399) than in continental UGC varieties (4,5% in texts, see Biagui 2018: 335, 339, 355), which may account for this difference. Note also that, in all Capeverdean ideophones with CCV syllables, this syllable type always appears in the first syllable of the word (37), although it can be repeated in reduplicated ideophones (38).

(37) Santiago *sakédu* ‘stand’ + *tran* (IDEO) > *sakédu tran* ‘stand **upright**’

(38) Santiago *gordu* ‘fat’ + *plátxi-plátxi* (IDEO) > *gordu plátxi-plátxi* ‘very fat, as fat as a pig’

Syllabic structure	No.	CCV. and CV. merged	No.	%	Syllabic type	
CCV.CCV	1	C(C)V.C(C)V	16	47%	open = C(C)V only	72%
CCV.CV	2					
CV.CV	13					
CCV.CV.CV	1	C(C)V.CV.CV	3.5	10%		
CV.CV.CV	2.5					
CCV.CV.CCV.CV	1	C(C)V.CV.C(C)V.CV	5	15%		
CV.CV.CV.CV	4					
CCVC	1	C(C)VC	4	12%	closed = C(C)VC only	24%
CVC	3					
CVC.CVC	4	CVC.CVC	4	12%		
CV.CVC	0.5	CV.CVC	0.5	1%	mixed	4%
VC.CV	1	VC.CV	1	3%		
<b>Total</b>	<b>34</b>	<b>Total</b>	<b>34</b>	<b>100%</b>		<b>100%</b>

Table 7: Syllabic structure of insular UGC ideophones (based on Santiago data):<sup>34</sup> syllabic types.

CCV. and CV. merged	No.	%	Syllables	No.	%
C(C)VC	4	12%	1	4	12%
C(C)V.CV	16	47%	2	20.5	63%
CV.CVC	0.5	1%			
CVC.CVC	4	12%			
VC.CV	1	3%			
C(C)V.CV.CV	3.5	10%	3	3.5	10%
C(C)V.CV.C(C)V.CV	5	15%	4	5	15%
<b>Total</b>	<b>34</b>	<b>100%</b>		<b>34</b>	<b>100%</b>

Table 8: Syllabic structure of insular UGC ideophones (based on Santiago data): number of syllables.

<sup>34</sup> As discussed in §6.3, Fogo ideophones are far less numerous than their Santiago counterparts.

CVC monosyllabic ideophones (4/34 items = 12%, including CCVC) and ideophones ending in a consonant (8,5/34 items = 25%) are much rarer in insular than in continental UGCs (68% and 99% respectively). This is largely due to the fact that insular UGCs admit a much lower number of consonants in the coda (Quint 2000a: 33-34; Moreira Tavares 2020: 75-76) than continental UGCs (Biagui 2018: 59-63). This contrast is particularly conspicuous for ideophonic roots shared by both UGC branches (39)-(40).

(39) Cacheu3 (continental)/Santiago (insular) *sai* ‘get out’

> Cacheu3 *sai fut* [fut] ‘get out **at once/quickly**’, see (4b)

≠ Santiago *sai futi* [ˈfuti], see (36)

*Fut* [fut] has a CVC structure and final [t], allowed by all continental UGCs, whereas *futi* [ˈfuti] ends in a paragogic [i] (and therefore has a CVCV structure), as final [t] is not allowed by Santiago insular UGC.

(40) CC *kumé* (continental)/Fogo (insular) *kumê* ‘eat’

> CC *kumé fep* [fep] ‘eat [sth.] **thoroughly**, eat **every last crumb** [of sth.]’

≠ Fogo *kumê fépu* [ˈfepu]

*Fep* [fep] has a CVC structure and final [p], allowed by all continental UGCs, whereas *fépu* [ˈfepu] ends in a paragogic [u] (and therefore has a CVCV structure), as final [p] is not allowed by Fogo insular UGC.<sup>35</sup>

### 6.1.3 Vowel patterns

A rule very similar to the one described for continental UGCs (§5.1.3) also applies to polysyllabic insular UGC ideophones: all vowels of an ideophonic root are identical (41), except for the post-tonic vowel (42), which is probably paragogic, see (39) and (40), and therefore not part of the root.

(41) Santiago *txera* ‘smell’ + *fututú* (IDEO) > *txera fututú* [futuˈtu] ‘have a **strong** smell’

<sup>35</sup> Our relatively reduced sample of insular ideophones seems to suggest that the choice of the paragogic vowel in insular UGCs depends on the preceding consonant (preC). If preC = [t, dʒ~ʒ, tʃ~c] (i.e. a dental or palatal consonant), the paragogic vowel is [i] (38). If preC = [b, f, p, k] (i.e. a labial or velar consonant), the paragogic vowel is [u] (39). The only known exception to these rules is Santiago *saki* [ˈsɛki] ~ *siki* [ˈsiki] (IDEO) > *sápa saki* ~ *sápa siki* ‘cut **clean**’.

(42) Santiago *labánta* ‘stand up’ + *fakati* (IDEO) > *labánta fakati* [fe'kɛti] ‘stand up **at once**’, compare with *fakat* in Cacheu (4a)

This rule applies to all five polysyllabic ideophonic roots attested in Santiago Capeverdean, all of which have cognates in continental varieties.

#### 6.1.4 Consonantic inventory

Only one known Santiago ideophone can be optionally pronounced with a voiced fricative (43).

(43) CC *intxi* ‘fill’ + *bipu* ~ *vipu* (IDEO) > *intxi bipu* or *intxi vipu* ‘fill [sth.] **up to the brim**’

Voiced fricatives are equally rare in Santiago core vocabulary (Quint 2000a: 112-114).

#### 6.1.5 Reduplication

While lexically reduplicated ideophones are rare in continental UGCs, they are much more frequent in Santiago Capeverdean. Out of a sample of 29 polysyllabic items (20.5 disyllables + 3.5 trisyllables + 5 quadrisyllables), nine insular UGC ideophones, i.e. close to one third of the total number of polysyllabic ideophones and  $9/34 = 26\%$  of all insular ideophones, have a fully (eight items, see (44)) or partially (one item, see (41)) reduplicated lexical base form. All quadrisyllabic Santiago ideophones, e.g. (44), are in fact reduplicated forms.

(44) Santiago *gordu* ‘fat’ + *bódji-bódji* (IDEO) > *gordu bódji-bódji* ‘**very fat, as fat as a pig**’

#### 6.1.6 Comparison of the phonology of insular and continental UGCs

Table 9 sums up the main commonalities and differences between continental and insular UGC ideophones. There is a clear phonological split between the two UGC branches and, in at least some domains (prosody, preferred syllabic structure, final – C, CCV), a greater Portuguese influence probably explains why insular UGCs differ

from continental UGCs. Indeed, Portuguese, like Capeverdean, is a stress language (prosody), with a dominant disyllabic word-pattern (preferred syllabic structure), few final consonants (final –C), and a relatively high frequency of CCV syllables (Quint 2000a: 35-40, Omnès 1988: 146). However, both insular and continental UGCs also share some features, one of which – the rarity of voiced fricatives – is clearly linked with an African common substrate (Quint 2000a: 112-114).

Upper Guinea Creoles			
Feature	Continental	Insular	Balance
Prosody (dominant pattern)	all high tone	stress	≠
Dominant syllable-type	CVC	CV	≠
Preferred syllabic structure	monosyllabic	disyllabic	≠
Final –C	pervasive	rare	≠
CCV	-	+	≠
Vowel harmony	+	+	=
Voiced fricatives	rare	rare	=
Lexical reduplication	< 10%	> 25%	≠

Table 9: Comparison of the phonology of insular and continental UGC ideophones.

## 6.2. Morphosyntax

### 6.2.1 Word order

Regarding word order, insular UGC ideophones behave exactly like their continental counterparts and immediately follow the element they modify, e.g. (38, 42), except when an object is inserted between a transitive verb and its associated ideophone (45, 46).

(45) Santiago

*e pegâ-m txápu na brásu*  
 s3SG seize.PFV-O1SG IDEO in/at arm

‘S/he seized **suddenly** my arm **with a firm grip** (lit. s/he seized me **firmly and suddenly** in [the] arm).’

(46) Santiago

*e abri pórtá uandan*  
 s3SG open.PFV door IDEO

‘S/he opened the door **wide**.’

### 6.2.2 Modified elements

As in continental UGCs, insular UGC ideophones usually associate with verbs. However, due to the fact that the distinction between verbs and adjectives is more clear-cut in Capeverdean than in continental UGCs,<sup>36</sup> seven Capeverdean ideophones (out of a total of 34, i.e. 21%) specifically associate with adjectives (as opposed to verbs).<sup>37</sup> Note that only *five* insular (Capeverdean) adjectives, namely *bedju* ‘old’, *bránku* ‘white’, *gordu* ‘fat’, *prétu* ‘black’ and *ségu* ‘blind’, have been found to associate with these *seven* ideophones. This difference in number is due to the fact that one and the same adjective can associate with two ideophones. To the best of our knowledge, there is no significant semantic difference between the two ideophones in (47).

- (47) Santiago *bedju* ‘old’  
> *bedju góbu-góbu* ‘very old’  
> *bedju kóti-kóti* ‘very old’

Furthermore, in Capeverdean, there is at least one case of a fossilized ideophone.

- (48) Santiago *sta* ‘be’ ~ *fika* ‘remain’ + *nunpriti* ‘naked’ > *sta/fika nunpriti* ‘be/remain naked’

Synchronically, *nunpriti* can be analysed as an adjective. However, it is quite probable that *nunpriti* can be further broken down into *nun* [nũ] ‘naked’ (a cognate of CC *nuŋ* (see (34)) < Port. *nu* ‘naked’) + *priti* [ˈpriti] (IDEO), a frozen ideophone intensifying the notion of nakedness. In other words, in continental UGCs, the reflex of the Portuguese adjective *nu* ‘naked’ was reanalysed as an ideophone, whereas in insular

---

<sup>36</sup> Continental *bran̄ku* ‘(be) white’, for example, is more appropriately described as a qualifier (or qualifying verb) which can combine with verbal morphemes and be used in the same slot as a verb, whereas Santiago (insular) *bránku* ‘white’ is not compatible with most verbal morphemes and better described as an adjective in the Portuguese (or Romance) meaning of the term. For more details about the question of adjectives and qualifiers in UGCs, see Biagui et al. (forthcoming), Quint (2000a: 297-300; 2000b: 103; 2023: 457, 459), Biagui (2018: 223-232), Biagui & Quint (2013: 46), Kihm (1994: 34-37, 148-149), Scantamburlo (1981: 44; 1999: 168-169) and Doneux & Rougé (1988: 27, 31, 39-40, 47, 49).

<sup>37</sup> We have classified Santiago Capeverdean *sakédu* ‘stand’ as a “qualifier” (§5.2.2), as, contrary to most Capeverdean adjectives but similar to its continental counterpart *sikidu* ‘stand’, it can behave as a verb when used in predicative function (Quint 2000a: 297-300).



UGCs, the reflex of Portuguese *nu* ‘naked’ kept its adjectival status but was intensified by a non-Romance ideophone, *priti*. We may therefore consider that, in Capeverdean, *nun* (in *nunpriti*) is a sixth instance of an adjective associating with an ideophone.

### 6.2.3 Morphology

In insular UGC ideophones, expressive reduplication is nonexistent while vowel lengthening and ideophone-to-verb derivation are marginally attested. The lengthening of stressed vowels is documented for color-associated ideophones and has both an attenuative/diminutive and laudatory value (49)-(50).

(49) Santiago *bránku* ‘white’ + *álbu* (IDEO)

> *bránku álbu* [ˈalbu] ‘**very white, as white as snow**’

> *bránku áálbu* [ˈa:lbu] ‘**so small, so nice and so white**’

(50) Santiago *prétu* ‘black’ + *finu* (IDEO)

> *prétu finu* [ˈfinu] ‘**very black, pitch-black**’

> *prétu fiinu* [ˈfi:nu] ‘**so small, so nice and so black**’

Some three cases of ideophone-to-verb derivation have been found (51)-(53).

(51) Santiago *sai* ‘get out’ + *futi* (IDEO)

> *sai futi* [ˈfuti] ‘get out **at once/quickly**’

> *futi* [ˈfuti] ‘get swiftly out [of a place], escape [from a place]’ (e.g. a goat from its pen or a bird from a trap...)

(52) Santiago *labánta* ‘stand up’ + *fakati* (IDEO)

> *labánta fakati* [feˈkɛti] ‘stand up **at once**’

> *fakatia* [feˈkɛtjɐ] ‘wriggle, move in all directions’

(53) Santiago *gordu* ‘fat’ + *plátxi-plátxi* (IDEO)

> *gordu plátxi-plátxi* [ˈplatʃi ˈplatʃi] ‘**very fat, as fat as a pig**’

> *plátxia* [ˈpletʃjɐ] ‘crash [to the ground]’

As all Santiago insular ideophones were checked for possible derived forms with a trained native-speaker, our list of deideophonic verbs is probably close to exhaustive. Contrary to CC, the derivational process is not morphologically regular: in (51), we have a conversion process, while in (52)-(53), a suffix [ɐ] is added to the ideophone. Note also that, in (53), the reduplication of the ideophonic base is lost in the derived verb. Furthermore, the meaning of the verb is not always predictable. In (52), it is the notion of SUDDEN MOVEMENT that seems to constitute the link between the ideophone and the derived verb, while the common semantic feature in (53) seems to be the SOFT, SHAPELESS CONSISTENCY of the fat (for the ideophone) and of something crashing to the ground and hence losing its original SHAPE and being somewhat SOFTENED (for the verb). These cases of ideophone-to-verb derivation are therefore highly idiosyncratic and contrast with the morphologically regular, semantically transparent ideophone-to-verb derivation that occurs in Casamance Creole.

#### 6.2.4 Comparison of the morphosyntactic properties of insular and continental UGCs

As shown in Table 10, the main morphosyntactic differences existing between continental and insular UGCs bear on their respective morphological properties. Generally speaking, continental UGCs display more diverse morphological patterns for their ideophones than insular UGCs.

Upper Guinea Creoles			
Feature	Continental	Insular	Balance
Word order	V-(O)-IDEO	V-(O)-IDEO	=
Associated elements	mostly verbs	mostly verbs (+ adjectives)	=
Expressive reduplication	+	-	≠
Vowel lengthening	intensive	attenuative	≠
Ideophone-to-verb derivation	regular	exceptional	≠

Table 10: Comparison of the morphosyntactic properties of insular and continental UGC ideophones.

### 6.3. Diatopic variation and inventory

As mentioned in §4.2, only two insular UGC varieties – Fogo and Santiago – have been investigated so far. The main difference between them concerns the size of their inventory: we have found only three ideophones in Fogo vs. 34 in Santiago, and all

Fogo ideophones have close equivalents in Santiago (Table 11). This result comes as no surprise as the African element has already been recognized as clearly less important in Fogo than in Santiago (Quint & Moreira Tavares 2019: 129-131). However, it must also be said that, as the vocabulary of Fogo has not been as thoroughly investigated as that of Santiago, we may have missed some Fogo ideophones.

Fogo			Santiago cognates
Modified element	Ideophone	Approximative meaning	
<i>sta</i> ‘be’ ~ <i>fikâ</i> ‘remain’	<i>enden</i> [ɛ̃ˈdɛ̃]	‘fully empty’	<i>iondon</i> [jõˈdõ] ~ <i>iandan</i> [jẽˈdẽ] ~ <i>uandan</i> [wẽˈdẽ]
<i>kumé</i> ‘eat’ ~ <i>bazâ</i> ‘spill’ ~ <i>panhâ</i> ‘catch/seize’ ~ <i>kunprâ</i> ‘buy’	<i>fépu</i> [ˈfɛpu]	‘completely, without leaving anything’	<i>fépu</i> [ˈfɛpu]
<i>prétu</i> ‘black’	<i>finu</i> [ˈfinu]	‘very’	<i>finu</i> [ˈfinu]

**Table 11:** The three known Fogo ideophones with their Santiago counterparts.<sup>38</sup>

It is also possible that more ideophonic items will be found on the two other islands of Sotavento (or Southern Capeverdean, see §2.1 and Map 2), namely Brava and Maio, which have not been explored for this matter. Furthermore, we expect that ideophones, which seem to be mostly a Niger-Congo feature in UGCs (see §5.4 and §6.4), are much less common in Barlavento (or Northern Capeverdean), where lexical African influence is weaker than in Sotavento (see §4.2).

## 6.4. Origin and comparison with continental UGCs

### 6.4.1 Origin of the insular UGC ideophones

Seventeen insular UGC ideophones (i.e. 17/34 = 50%) can be traced back to Mandinka and/or Wolof (Table 12). Basically, the ratio of Mandinka and Wolof is similar to what has been observed in continental UGCs (see §5.4 and Table 6), and most of the African-derived insular UGC ideophones have cognates in continental

<sup>38</sup> For Fogo, *enden* is attested in Moreira Tavares (2020: 237), *fépu* in Moreira Tavares (2020: 273) and Quint & Moreira Tavares (2019: 152), and *finu* in Quint & Moreira Tavares (2019: 148).

UGCs. If we consider the distribution types (with due caution because of the limited amount of data), Mandinka seems to predominate in the ‘common core’ type while undoubtedly Wolof-derived ideophones appear only in the ‘CC-only’ and ‘minority models’ types. Here again, the distribution profile is very similar to what has been observed for continental UGC African-derived ideophones (ADI).

Origin					
Distribution type <sup>39</sup>	Wolof	Mandinka	Wolof or Mandinka	Total ADI	Continental cognates
CC-only	1	0	0	1	1
Minority models	2	3	1	6	4
Common core	0	3	3	6	5
Other distribution	0	1	2	3	2
NA <sup>40</sup>	1	0	0	1	1
<b>Total</b>	<b>4</b>	<b>7</b>	<b>6</b>	<b>17</b>	<b>13</b>

**Table 12:** African-derived insular UGC ideophones according to their distribution type as defined in Table 4.

Like on the continent, some insular UGC ideophones come from Portuguese (54) or can be derived both from an African language and/or Portuguese (55).

(54) Santiago *bránku* ‘white’ + *álbu* (IDEO) > *bránku álbu* ‘**very** white, **as** white **as snow**’, with *álbu* < Portuguese *alvo* ‘bright/intense white’

(55) Santiago *prétu* ‘black’ + *finu* (IDEO) > *prétu finu* ‘**very** black, **pitch**-black’, with *finu* < Mandinka *fɪŋ* ‘be black’ and/or Portuguese *fino* ‘fine, refined’

<sup>39</sup> The distribution types are those defined for continental UGCs, i.e. if a given insular UGC ideophone X is the cognate or semantic equivalent of a continental ideophone Y, X is ascribed to the distribution type Y belongs to.

<sup>40</sup> Some insular ideophones have no attested (NA) cognate or semantic equivalent in the list of 81 Casamance Creole (CC) ideophones which has been used as the standard for continental UGCs (§4.1). However, in some cases, we were able to find afterwards a continental cognate or semantic equivalent outside the CC list, e.g. Santiago *pila* ‘crush’ ~ *kebra* ‘break’ + *moku* (IDEO) > *pila/kebra moku* ‘break/crush **completely**’. After *moku* was found in Santiago Capeverdean, it was also found with a similar meaning and form both in Casamance and Bissau3 continental varieties. Furthermore, *moku* can be traced to a Wolof term, *mokk* [mɔːkkə] ‘be ground/crushed’ (Diouf 2003: 226; Dieng 1985: 247). But as *moku* was not included in the initial list of 81 CC ideophones, it is counted in Table 11 as an insular African-derived ideophone belonging to the ‘non-attested (NA)’ type.

#### 6.4.2 Comparison with continental UGCs

The first difference between insular and continental UGC ideophones concerns their number: there are less ideophones on the islands than on the continent. Only 31 ideophones<sup>41</sup> from insular varieties correspond to the 81 ideophones of the standard Casamance Creole list. Put differently, 50 ideophones of the standard CC list lack an equivalent in Santiago Capeverdean, the only insular variety to have been extensively checked. The number of missing equivalents is thus much higher than for all continental varieties that have been checked with consultants, where there are between 23 and 36 gaps per list (see §5.3.2 and Table 5).

Admittedly, the standard CC list is probably far from exhaustive, and there must be many more ideophones, both in continental and insular varieties, that were not taken into account in the present study. However, from our own experience of both UGC branches (including daily practice thereof), we can say quite confidently that the ideophone category is less developed in insular than in continental varieties. As the category seems to draw most of its members from the African substrate and adstrate (see 5.4.), which is stronger on the continent than on the archipelago of Cape Verde, the difference in the number of ideophones seems to be consistent with what we know about the history of UGCs.

When we take the distribution types defined for continental UGC ideophones (see §5.3.1 and Table 4) and apply them to three subsets of our insular UGC sample of ideophones (Table 13), namely (i) the semantic equivalents of CC ideophones, (ii) the cognates of CC ideophones and (iii) the African-derived ideophones (ADI), the most striking observation is the similarity between continental and insular varieties regarding the ratio of each distribution type, no matter which subset is taken into account. For all subsets, the items belonging to the continental ‘common core’ type make up approximately one half of the total number of insular ideophones, just as in CC. All in all, insular and continental UGC ideophones share the same distribution profile.

---

<sup>41</sup> We have excluded the three insular ideophones belonging to the ‘NA’ type (see fn. 40), as they are precisely defined by the fact that they have no CC equivalent. Hence the value of 31 insular UGC ideophones (= 34 (total) – 3 (NA)) which we consider for the comparison with continental UGCs.

Distribution type	Insular (Santiago) ideophones			
	CC	Semantic equivalents	Continental cognates	ADI
CC-only	22	4	3	1
Minority models	14	7	5	6
Common core	<b>39</b>	<b>15</b>	<b>9</b>	<b>6</b>
Other distribution	9	5	3	3
NA	0	3	1	1
Total	84	34	21	17

**Table 13:** Comparison of continental (CC) UGC ideophones with different subsets of insular UGC ideophones according to their distribution type as defined in Table 4.

However, if we consider the semantic types of insular UGC ideophones (Table 14), a difference appears between the two UGC branches. Regarding the total sample of insular UGC ideophones, the semantic distribution is quite comparable to what has been observed for continental varieties, with a clear predominance of intensifying ideophones: 25 out of 34, i.e. 74%, a value equivalent to what was found in the CC-standard list (75%, see §4.3). Nonetheless, if we take into account, for each semantic type, the number of insular UGC ideophones that have a plausible cognate in continental UGCs, a clear split appears:

- Almost all (8/9) insular UGC non-intensifying ideophones have cognates in continental UGCs.
- Only one half (13/25) of insular UGC intensifying ideophones have cognates in continental UGCs.

In other terms, insular intensifying ideophones seem to be less easily traceable to proto-UGC than ideophones of other semantic categories. Yet it appears that it is precisely this semantic category that was favoured and considered as the most prototypical during the elaboration of the standard CC list (§4.3). Why then are intensifying ideophones less prone to be shared between continental and insular UGCs? The main underlying reason may well be once more the respective degree of Niger-Congo influence. As a matter of fact, intensifying ideophones are certainly a salient semantic category when contrasting Niger-Congo languages with Portuguese, the main lexifier of both insular and continental UGCs (§2.2). Whereas, for instance, sound ideophones (or “onomatopoeias”) are found in Portuguese (and in other Romance languages), a word category matching Niger-Congo intensifying ideophones

is missing. This may well explain the relative lack of stability of intensifying ideophonic roots in insular UGCs: these varieties, submitted to a stronger Portuguese influence, must have aligned their semantic organization with their Romance lexifier and lost African features (such as the use of intensifying ideophones) that contrasted too drastically with Portuguese. In contrast, continental UGCs, which had less intense contacts with Portuguese (and more African influence), were more prone to retain Niger-Congo material.

Semantic type		Total	Continental cognates
Intensifying		25	13
Non-intensifying		9	8
of which:	Sound (onomatopoeia)	4	4
	Immediate	3	3
	Others	2	1
Total		34	21

Table 14: Semantic types of insular UGC ideophones.

The semantic field of color provides a good example of the difference between insular and continental ideophones regarding intensifying ideophones. Many Niger-Congo languages make a lexical distinction between three (more rarely four, e.g. Koalib) basic colors (‘white’ = FULL LIGHT, ‘red’ = FULL COLOR and ‘black’ = NO LIGHT, NO COLOR, see Segerer & Vanhove (2019: 292, 315-316); Rougé (1988: 40-41)), each of which is intensified by a specific ideophone (Table 15). These color intensifiers are among the most stable ideophones, and it is possible to reconstruct proto-forms for some branches of Niger-Congo (for Kordofanian, see examples in Quint 2018: 202-204).

Intensifying ideophone						
Color	Niger-Congo languages				UGCs	
	Mandinka	Wolof	Djifanghor Nyun	Koalib	Proto-conti- nental UGC	Insular
‘white’	<i>fér</i> [+léw ~ wéj]	<i>fur</i> ~ <i>tàll</i>	<i>pat</i>	<i>ppéppèr</i>	*/nok/	[‘albu]
‘red’	<i>cáw</i> ~ <i>táw</i> ~ <i>táráw</i>	<i>coy</i> ~ <i>cur(r)</i>	<i>úutut</i>	<i>ccél-ccél</i> ~ <i>ccéccél</i>	*/wak/	—
‘black’	<i>kím</i> ~ <i>kírím</i> [+hót ~ nót]	<i>kuk(k)</i>	<i>ñotótót</i>	<i>ťíťít</i>	*/fandan/	[‘finu]

Table 15: Intensifying ideophones for the three basic colours ‘white, red, black’ in several Niger-Congo languages, in proto-continental and in insular UGCs.

Table 15 shows that the system of three color-intensifying ideophones of continental UGCs fits in with the general Niger-Congo semantic pattern. In contrast, insular UGCs differ from the Niger-Congo (and continental UGC) model in two respects: They lack an intensifier for ‘red’, and they have developed Romance (or partially Romance)-derived ideophones — see (54) and (55) – to intensify ‘black’ and ‘white’, whereas continental UGCs have African (or at least non-Romance) intensifying ideophones for the same colors.

In sum, the semantic field of color illustrates quite well the difference between the system of continental UGC ideophones, which has stronger semantic and phonological links with Niger-Congo languages, and the system of insular UGC ideophones, whose African-derived component, albeit still visible, is weaker, both phonologically and semantically, than in continental UGCs.

#### 6.4.3 Comparison of the origin and semantic behavior of insular and continental UGC ideophones

Table 16 summarizes the main commonalities and differences between continental and insular UGC ideophones regarding their origin and semantic organization.

Upper Guinea Creoles			
Feature	Continental	Insular	Balance
Origin	Mandinka + Wolof	Mandinka + Wolof	=
Inventory	> 40	< 40	≠
Dominant distribution type	common core	common core	=
Intensifying ideophones: cognates with continental UGCs	100%	50%	≠

**Table 16:** Comparison of the origin and semantic organization of insular (= Capeverdean) and continental UGC ideophones.

## 7. Conclusion

This paper has described in great detail the phonology, morphosyntax, diatopic variation, origin and semantic properties of UGC ideophones. It has also provided a systematic comparison of the characteristics of this particular word class in two branches of UGCs: (i) continental (Casamance and Guinea-Bissau) and (ii) insular



(Capeverdean) UGCs. In our study, we were able to identify and define several tens of ideophones in continental and insular UGCs. The existence of these ideophones seems to be mainly due to the influence exerted by the Niger-Congo African substrate (for both insular and continental UGCs) or/and adstrate (continental UGCs only). The observation that many ideophonic roots are shared by continental and insular UGCs reinforces the hypothesis of a common stage (the proto-Upper-Guinea-Creole or proto-UGC) from which both UGC branches are descended. The fact that various African-derived ideophones can be traced back to proto-UGC implies that this common stage must have been different from Portuguese (§2.2), a Romance language which does not have an ideophone category. Yet, despite obvious resemblances, continental and insular UGC ideophones also reflect a clear split between the two branches, as the influence of Niger-Congo languages is significantly heavier on the continent. This important result is confirmed by other similar findings (Rougé 1999, Quint 2000b: 99-117, Quint & Moreira Tavares 2019: 126-127) regarding the respective influence of the African substrate on insular and continental UGCs.

In the future, the present research could be expanded in several directions. Increasing the sample of ideophones would improve our knowledge of all continental and insular varieties. In particular, the three Sotavento (Southern Capeverdean) varieties other than Santiago should be thoroughly investigated, as we have almost no data for Brava and Maio and too few for Fogo. Furthermore, our assessment of the substrate and/or adstrate influence would be strengthened if more Niger-Congo languages were checked. Balant (bjt/ble), Jola, Manjak (mfv), Papel (pbo) and Nyun, which are all spoken in direct contact with one or several of today's continental UGCs, probably have a lot to offer in this respect. Finally, a refinement and an elaboration of the semantic and distribution classifications that have been proposed in this paper seems desirable. Large, accessible, searchable corpora for the UGCs would allow us to pay more attention to the collocations involving ideophones and their actual usage in spontaneous speech.

It is clear that the present study is far from having exhausted the question of UGC ideophones, which remain a most exciting topic for anyone interested in language contact effects on ideophones, in general, and the grammar and genesis of the Afro-Portuguese varieties of Cape Verde, Casamance and Guinea-Bissau, in particular.

## Acknowledgements

We would like to thank Aires Semedo and Karina Moreira Tavares for their comments and advice on Capeverdean data, Braima Sam Mendes for his help during our fieldwork in Guinea-Bissau, and Jean-Baptiste Maistre for his bibliographical support. Our thanks also go to all the native speakers who were kind enough to share with us their knowledge of their respective varieties. We also are grateful to the anonymous reviewers of this paper for their helpful suggestions and to special issue editors for their meticulous editing.

## Abbreviations

1 = 1 <sup>st</sup> person	IDEO/IDEO = ideophone	PL = plural
2 = 2 <sup>nd</sup> person	IPFV = imperfective	RED = reduplicated
3 = 3 <sup>rd</sup> person	NA = non-attested	S = stressed
ADI = African-derived ideophone	NEG = negation	S/S = subject
ASS = assertive	O/O = object	SG = singular
CC = Casamance Creole	PFV = perfective	SUB = subordinator
DEM = demonstrative	Port. = Portuguese	U = unstressed
F = feminine	preC = preceding consonant	UGC = Upper Guinea Creole
H = high tone	PROG = progressive	
HAB = habitual	PROX = proximal	

## References

- Biagui, Noël B., Jean-François Nunez & Nicolas Quint. (forthcoming). *Casamance Creole*. In Lüpke, Friederike (ed.), *Oxford guide to the world's languages: Atlantic*. Oxford: Oxford University Press.
- Biagui, Noël B. & Nicolas Quint. 2013. *Casamance Creole*. In Susanne Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *Survey of pidgin and creole languages*, Vol. 2: *French-based, Portuguese-based and Spanish-based languages*, 40–49. Oxford: Oxford University Press.
- Biagui, Noël B. 2018. *Description générale du créole afro-portugais de Ziguinchor (Sénégal)*. Paris: Karthala.
- Biagui, Noël B. (in preparation). *Dictionnaire créole casamançais-français*.
- Childs, Tucker. 1994. Expressiveness in contact situations: The fate of African Ideophones. *Journal of Pidgin and Creole Languages* 9(2). 257–282.

- Couto, Hildo H. 1994. *O Crioulo Português da Guiné-Bissau*. Hamburg: Helmut Buske.
- Couto, Hildo H. 1995. Exclusive particles (ideophones) in Guinea-Bissau creole. In Philip Baker (ed.), *From contact to creole and beyond*, 207–215. London: University of Westminster Press.
- Creissels, Denis. 1994. *Aperçu sur les structures phonologiques des langues négro-africaines*. Grenoble: ELLUG.
- Creissels, Denis. 2006. *Syntaxe structurale, une introduction typologique 1. Catégories et Construction*. Paris: Lavoisier.
- Creissels, Denis. 2012. *Lexique mandinka-français*. Unpublished MS. Available online at: [http://www.deniscreissels.fr/public/Creissels-lexique\\_mandinka\\_2012.pdf](http://www.deniscreissels.fr/public/Creissels-lexique_mandinka_2012.pdf) (Accessed 28/03/2023).
- Creissels, Denis. 2013. *Le mandinka – phonologie, grammaire, textes*. Paris: Karthala.
- Creissels, Denis, Sidia Jatta & Jobarteh Kalifa. 1982. *Lexique mandinka-français. Mandenkan 3*.
- Dalphinis, Morgan. 1986. *African language influences in creoles lexically based on Portuguese, English and French, with special reference to Casamance Kriul [sic], Gambian Krio and St. Lucian Patwa*, Vol. 1. Doctoral dissertation, University of London.
- Dieng, Sady. 1985. *Étude contrastive des lexies du créole endogène “portugais” et du wolof*. Doctoral dissertation, Université Lumière-Lyon 2.
- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass* 6/10. 654–672.
- Dingemanse, Mark. 2019. ‘Ideophone’ as a comparative concept. In Kimi Akita & Prashant Pardeshi (eds.), *Ideophones, mimetics and expressives*, 13–34. Amsterdam: John Benjamins.
- Diouf, Jean-Léopold. 2003. *Dictionnaire wolof-français et français-wolof*. Paris: Karthala.
- Doneux, Jean-Léonce & Jean-Louis Rougé. 1988. *En apprenant le créole à Bissau ou Ziguinchor*. Paris: L’Harmattan.
- Fal, Arame, Rosine Santos & Jean-Léonce Doneux. 1990. *Dictionnaire wolof-français, suivi d’un index français-wolof*. Paris: Karthala.
- Jacobs, Bart. 2012. *Origins of a creole. The history of Papiamentu and its African ties*. Berlin: De Gruyter.
- Kihm, Alain. 1994. *Kriyol syntax: The Portuguese-based creole language of Guinea-Bissau*. Amsterdam: John Benjamins.

- Mbodj, Chérif. 1979. *Phonologie du créole de Guinée-Bissau*. Dakar: Centre de linguistique appliquée de Dakar.
- Mbodj, Chérif. 1984. *Grammaire du créole de Guinée-Bissau*. Bissau: Unpublished MS.
- Moreira Tavares & Ana Karina. 2020. *Documentation et description grammaticale et lexicale du créole afro-portugais de l'île de Fogo (République du Cap-Vert, Afrique de l'Ouest) / Documentação e descrição gramatical e lexical do crioulo afro-português da ilha do Fogo (República de Cabo Verde, África Ocidental)*. Doctoral dissertation, IN-ALCO, Paris.
- Nuckolls, Janis B., Joseph A. Stanley, Elizabeth Nielsen & Roseanna Hopper. 2016. The systematic stretching and contracting of ideophonic phonology in Pastaza Quichua. *International Journal of American Linguistics* 82(1). 95–116.
- Omnès, Robert. 1988. Les structures syllabiques du castillan, étude synchronique et diachronique, comparaison avec 28 autres langues. In *Actes du 3e colloque de linguistique hispanique*, 135–155. Grenoble: Université Stendhal-Grenoble 3.
- Quint, Nicolas. 1998. *Dicionário de caboverdiano-português*. Lisbon: Verbalis.
- Quint, Nicolas. 1999. *Dictionnaire cap-verdien-français*. Paris: L'Harmattan.
- Quint, Nicolas. 2000a. *Grammaire de la langue capverdienne (description diachronique et synchronique du créole afro-portugais des îles du Cap-Vert)*. Paris: L'Harmattan.
- Quint, Nicolas. 2000b. *Le cap-verdien: origines et devenir d'une langue métisse*. Paris: L'Harmattan.
- Quint, Nicolas. 2006. *Phonologie de la langue koalibe*. Paris: L'Harmattan.
- Quint, Nicolas. 2008. *L'élément africain dans la langue capverdienne / Africanismos na língua caboverdiana*. Paris: L'Harmattan.
- Quint, Nicolas. 2009a. *The phonology of Koalib, a Kordofanian language from the Nuba Mountains (Sudan)*. Cologne: Rüdiger Köppe.
- Quint, Nicolas. 2009b. O cabo-verdiano: uma língua mundial. *Revista de Estudos Caboverdianos* 3. 129–144.
- Quint, Nicolas. 2018. Koalib ideophones: a first overview of their main characteristics. In Gertrud Schneider-Blum, Birgit Hellwig & Gerrit J. Dimmendaal (eds.), *Nuba Mountain languages studies: new insights*, 175–207. Cologne: Rüdiger Köppe.
- Quint, Nicolas. 2023. Guinea-Bissau. In Ursula Reutner (ed.), *Manual of Romance languages in Africa*, 447–465. Berlin: De Gruyter.
- Quint, Nicolas & Karina Moreira Tavares. 2019. The common African lexical core of Upper Guinea Creoles and its historical significance. *Journal of Ibero-Romance Creoles* 9. 115–161.

- Rougé, Jean-Louis. 1988. *Petit dictionnaire étymologique du kriol de Guinée-Bissau et de Casamance*. Bissau: Instituto Nacional de Estudos e Pesquisa.
- Rougé, Jean-Louis. 1999. Apontamentos sobre o léxico de origem africana dos crioulos da Guiné e de Cabo Verde (Santiago). In Klaus Zimmermann (ed.), *Lenguas criollas de base lexical española y portuguesa*, 49–67. Frankfurt am Main: Vervuert.
- Rougé, Jean-Louis. 2004. *Dictionnaire étymologique des créoles portugais d’Afrique*. Paris: Karthala.
- Scantamburlo, Luigi. 1981. *Gramática e dicionário da língua criol da Guiné-Bissau (GCr)*. Bologna: Editrice Missionaria Italiana.
- Scantamburlo, Luigi. 1999. *Dicionário guineense*, Vol. 1: *Introdução e notas gramaticais*. Lisbon: Edições Colibri/Faspebi (Fundação para o Apoio ao Desenvolvimento dos Povos do Arquipélago de Bijagós).
- Scantamburlo, Luigi. 2002. *Dicionário do guineense*, Vol. 2: *Dicionário guineense-português. Disionariu guinensi-purtuguis*. Lisbon: Colibri/Faspebi.
- Segerer, Guillaume & Martine Vanhove. 2019. Color naming in Africa. In Ida Raffaelli, Daniela Katunar & Barbara Kerovec (eds.), *Lexicalization patterns in colour naming: a cross-linguistic perspective*, 287–330. Amsterdam: John Benjamins.
- Truppi, Chiara & Patrícia Costa. 2019. *On ideophones in Portuguese Creoles*, paper presented at *Societas Linguistica Europaea – 52<sup>nd</sup> Annual Meeting*, 21<sup>st</sup>–24<sup>th</sup> August 2019 in Leipzig.
- Wilson, William A. A. 1962. *The Crioulo of Guiné*. Johannesburg: Witwatersrand University Press.

## CONTACT

nicolas.quint@cnrs.fr

biaguinoelbernard@yahoo.fr

## Appendices

## Appendix 1: Continental UGC ideophones classified according to their distribution type

Table 1.1: ‘CC-only’ type.

		CC ideophone				Associated element			Other continental UGC equivalents
No.	Ideophone	SS	Etymon	SEM	RED	Form <sup>42</sup>	English	PoS	
(1)	<i>bok-bok</i>	CVC-CVC	—	intens.	Yes	<i>termé</i>	‘shiver’	v.	—
(2)	<i>ca</i>	CV	—	intens.	No	<i>fritá</i>	‘fry’	v.	—
(3)	<i>cañar</i>	CVCVC	—	intens.	No	<i>jugtá ~ jugutá</i>	‘jump’	v.	<i>caŋ</i> (+ <i>saltá</i> ‘jump’) (Cacheu3)
(4)	<i>cip</i>	CVC	—	intens.	No	<i>disí</i>	‘get down’	v.	<i>cem ~ caŋ</i> (Cacheu2)
(5)	<i>colop</i>	CVCVC	Mandinka©	intens.	No	<i>jugtá ~ jugutá</i>	‘jump’	v.	<i>cas</i> (Cacheu1 + Cacheu3); <i>cem ~ caŋ</i> (Cacheu2)
(6)	<i>fes</i>	CVC	—	intens.	No	<i>simí</i>	‘disappear’	v.	—
(7)	<i>fututut</i>	CVCVCVC	—	intens.	Yes	<i>kerá</i>	‘smell’	v.	—
(8)	<i>kɛc</i>	CVC	—	intens.	No	<i>jundá</i>	‘drink’ <sup>43</sup>	v.	<i>forgot</i> (Cacheu3)
(9)	<i>kɛkɛt</i>	CVCVC	—	intens.	Yes	<i>risu</i>	‘be hard/tough’	qual.	—

<sup>42</sup> As Casamance Creole is the standard for the comparison of continental UGCs (§2.1), the form of the associated element is the one attested in CC, unless otherwise specified.

<sup>43</sup> The basic meaning of *jundá* is ‘pull’. However, *kis* intensifies a figurative meaning of *jundá*, namely ‘drink (alcohol)’.

		CC ideophone				Associated element			Other continental UGC equivalents
No.	Ideophone	SS	Etymon	SEM	RED	Form <sup>42</sup>	English	PoS	
(10)	<i>kamaɟ</i>	CVCVC	Wolof©	immdt.	No	<i>pagá</i>	‘switch off’	v.	<i>fup</i> (?) (Cacheu3); <i>vut</i> (Bissau1)
(11)	<i>kik</i>	CVC	—	intens.	No	<i>sikidu</i>	‘be straight’ (pole/stake)	qual.	—
(12)	<i>fof</i>	CVC	Mandinka©	intens.	No	<i>lebi</i>	‘be light’	qual.	<i>kef</i> (Cacheu3)
(13)	<i>koŋ</i>	CVC	Wolof©	intens.	No	<i>risu, seku</i>	‘be hard/tough, be dry’	qual.	—
(14)	<i>mɛp</i>	CVC	—	intens.	No	<i>barí</i>	‘get, win, collect’	v.	<i>fep</i> (Cacheu2); <i>bup</i> (Cacheu3)
(15)	<i>mes</i>	CVC	Wolof©	intens.	No	<i>perdé</i>	‘disappear’	v.	<i>buk</i> (Cacheu3)
(16)	<i>pejeɟ</i>	CVCVC	—	speech (‘clearly’)	No	<i>kontá</i>	‘tell’	v.	—
(17)	<i>paraw</i>	CVCVC	—	intens.	No	<i>camí</i>	‘get drunk’	v.	<i>tɔɾɔk</i> (Cacheu2)
(18)	<i>pip</i>	CVC	—	speech (‘threateningly’)	No	<i>falá</i>	‘say’	v.	—
(19)	<i>pul</i>	CVC	Wolof©	immdt.	No	<i>sey</i>	‘get out’	v.	<i>fas</i> (Cacheu2); <i>fat</i> (Bissau1)
(20)	<i>rat</i>	CVC	—	intens.	No	<i>dá</i>	‘drink’	v.	—
(21)	<i>tɛkɛp</i>	CVCVC	—	intens.	No	<i>pañá</i>	‘thicken’	v.	—
(22)	<i>was</i>	CVC	Mandinka + + Wolof©	intens.	No	<i>pajigá</i>	‘scatter’	v.	<i>buk</i> (Cacheu3); <i>fep</i> (Bissau1)

Table 1.2: ‘minority model’ type.

No.	Distr. profile	Continental UGC common ideophone					Associated element			Other continental UGC equivalents	
		Ideophone	Attested forms	SS	Etymon	SEM	RED	Form	Meaning		PoS
(1)	CC+2	*/bap/	<i>bap</i> (CC); <i>vap</i> (Bissau1); <i>wap</i> (Bissau3)	CVC	Wolof©	intens. + immdt.	No	<i>sintá/kay</i>	‘sit, fall’	v.	<i>rip</i> (Cacheu3)
(2)	CC+2	*/fas/	<i>fas</i> (CC+Cacheu1); <i>fas</i> (?) ~ <i>fes</i> (+ <i>rapá</i> ‘shave’) ~ <i>feles</i> (Geba)	CVC	Wolof©	intens.	No	<i>limpu</i>	‘be clean’	qual.	<i>wañja(η)</i> ~ <i>wac</i> (?) (Geba)
(3)	CC+3	*/fuf/ ~ */fus/	<i>fuf</i> (CC+Bissau3); <i>fus</i> (Cacheu2+Bissau1)	CVC	Wolof©	intens.	No	<i>mas</i>	‘be more’	v.	—
(4)	CC+1	*/kef/	<i>kef</i> (CC+Cacheu1)	CVC	—	intens.	No	<i>fartá</i>	‘be satisfied’ (eating)	v.	<i>diŋ</i> ~ <i>dir</i> (Geba); <i>kuŋ</i> (Geba+Cacheu2); <sup>44</sup> <i>lip</i> (Cacheu3)
(5)	CC+3	*/kokorot/	<i>kokorot</i> (CC+Cacheu1+ Cacheu3+Bissau1)	CVCVCVC	—	intens.	Yes	<i>beju</i>	‘be old’	qual.	—
(6)	CC+2	*/las/	<i>las</i> (CC+Bissau1); <i>lac</i> (Cacheu1)	CVC	Wolof©	intens.	No	<i>ditá</i>	‘lie’	v.	<i>caŋ</i> (+ <i>lastrá</i> ‘stretch out’) (Geba)
(7)	CC+1	*/net/	<i>net</i> (CC+CA2)	CVC	—	intens.	No	<i>pí</i>	‘put’	v.	—
(8)	CC+1	*/pirkit/	<i>pirkit</i> (CC+Geba)	CVCCVC	Mandinka©	immdt.	No	<i>labantá</i>	‘stand up’	v.	—
(9)	CC+3	*/puf/	<i>puf</i> (CC+Geba+Cacheu3+)	CVC	—	intens.	No	<i>furá</i>	‘make a hole,	v.	<i>fut</i> (Cacheu3)

<sup>44</sup> As Geba and Cacheu2 have the same form (*kuŋ*), we could have considered this example as an illustration of the CC vs. Guinea-Bissau subtype of the ‘other distribution’ type (§3.3.1 and Table 1.4). However, the ideophone \*/kuŋ/ is also attested in at least seven different continental UGCs as intensifying *yĩncí* ‘be full’ (Table 1.3). Hence, it seems that the form *kuŋ* attested here in Geba and Cacheu2 is due to a confusion between being ‘full’ and ‘satisfied’ (= ‘full of food’) rather than to a genuine semantic peculiarity shared by Geba and Cacheu2.



Continental UGC common ideophone							Associated element			Other continental UGC equivalents	
No.	Distr. profile	Ideophone	Attested forms	SS	Etymon	SEM	RED	Form	Meaning	PoS	
			Bissau1)						‘pierce’		
(10)	CC+3	*/tar/	<i>tar</i> (CC+Geba+Cacheu3+ Bissau1); <i>tax</i> [tax] (Geba); <i>par</i> (Cacheu1)	CVC	—	intens. + onom.	No	<i>bafatiyá</i>	‘slap’	v.	<i>baw</i> (Geba); <i>paw</i> (Cacheu1 + Cacheu2) <sup>45</sup> ; <i>fak</i> (Cacheu1); <i>fap</i> (Bissau3); <i>vap</i> (Cacheu3)
(11)	CC+3	*/tep/	<i>tep</i> (CC+ +Bissau3); <i>tep</i> (+ <i>fartá</i> ‘be satisfied’) (Cacheu1); <i>tip</i> ((+ <i>fartá</i> ) (Cacheu3)	CVC	Mandinka©	intens.	No	<i>yĩncí</i>	‘fill’ ~ ‘be full’	v.	—
(12)	CC+1	*/tim/	<i>tim</i> (CC+Geba)	CVC	—	onom. (‘hard surface’)	No	<i>kay</i>	‘fall’	v.	<i>fat</i> (Cacheu1); <i>kaŋkalaŋ</i> ~ <i>puf</i> (Cacheu2); <i>tuc</i> (Cacheu3); <i>tup</i> (Bissau1)
(13)	CC+2	*/wac/	<i>wac</i> (CC); <i>wap</i> (Cacheu1); <i>vac</i> (+ <i>yabrí</i> ‘open’) (Bissau1)	CVC	Mandinka©	intens.	No	<i>fendé</i>	‘split’	v.	<i>parac</i> ~ <i>típ</i> (+ <i>yabrí</i> ‘open’) (Bissau1)
(14)	CC+2	*/jambalaŋ/ ~ */wandal aŋ/	<i>yambalaŋ</i> (CC); <i>wandalaŋ</i> (Cacheu1 + Cacheu2)	CVCCVCVC	Wolof©	visual (‘desert’)	No	<i>sá</i> (CC) (= <i>sta</i> )/ <i>fiká</i>	‘be’	v.	—

<sup>45</sup> Here also, as for \*/kef/ above, we have come to the conclusion that the common form *paw* found in Cacheu1 and Cacheu2 does not necessarily illustrate a semantic peculiarity shared by these two varieties. The sound produced by a slap is probably \*/tar/ across most of the UGC continental varieties, but some consultants resorted to other, less specialized sound ideophones such as *paw*, which we have found in other varieties in quite different contexts (e.g. to describe the noise produced by a broken stick in CC).

Table 1.3: ‘common core’ type.

Continental UGC common ideophone											Other CUGC equiv.
No.	Distr. profile	Ideophone	Attested forms	SS	Etymon	SEM	RED	Form	Meaning	PoS	
(1)	CC + 7	*/baw/ <sup>46</sup>	<i>baw</i> (CC + Geba + Cacheu1 + Cacheu2 + Bissau2); <i>badaw</i> (Cacheu3 + Bissau1 + Bissau3); <i>bak</i> (Bissau2)	CVC	—	intens.	No	<i>forti</i>	‘be sour’	qual.	—
(2)	CC + 4	*/bik/	<i>bik</i> (CC + Geba + Cacheu1 + Bissau1)	CVC	—	intens.	No	<i>kabá</i>	‘be finished’	v.	—
(3)	CC + 6	*/buk/	<i>buk</i> (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	—	intens.	No	<i>kabá, pajigá</i>	‘be finished, scatter’	v.	—
(4)	CC + 4	*/bup/	<i>bup</i> (CC + Bissau1); <i>vup</i> (Geba); <i>bip</i> (Cacheu1); <i>vip</i> (Cacheu1 + Cacheu3)	CVC	—	onom.	No	<i>kay</i>	‘fall’	v.	—
(5)	CC + 7	*/caŋ/ ~ */kaŋ/ <sup>47</sup>	<i>caŋ</i> (Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau2 + Bissau3); <i>kaŋ</i> (CC + Cacheu1 + Cacheu3 + Bissau1); <i>taŋ</i> (Bissau3)	CVC	—	intens.	No	<i>firmá/sikidu</i>	‘stand’ (person)	v./qual.	—
(6)	CC + 4	*/cap/	<i>cap</i> (CC + Geba + Cacheu1 + Cacheu3 + Bissau1) ~ <i>ñap</i> [ɲap] (Geba)	CVC	Mandinka©	intens.	No	<i>pañá/pegá</i>	‘catch’	v.	—
(7)	CC + 4	*/corot/	<i>corot</i> (CC + Cacheu1); <i>corot</i> (+ <i>pikininu</i> ‘be small’) (Geba + Cacheu2) <i>ñeret</i>	CVCVC	Mandinka©	atten.	No	<i>pañá, partí</i>	‘take, give’	v.	—

<sup>46</sup> The form \*/baw/ has been retained as the base form for this ideophone against *badaw*, which is almost certainly an extended (expressive) form of *baw*.

<sup>47</sup> The form \*/kaŋ/ is obviously the same as the one that is used to intensify *tesu* ‘be solid’ (see Table 1.4). Some continental UGC varieties seem to use the same form to intensify ‘be solid’ and ‘stand’, while others use one specific ideophone for each of these two meanings.

Continental UGC common ideophone										Associated element	Other CUGC equiv.
No.	Distr. profile	Ideophone	Attested forms	SS	Etymon	SEM	RED	Form	Meaning	PoS	
			(+ <i>pikininu</i> ) (Cacheu3)								
(8)	CC + 5	*/culup/	<i>culup</i> (Geba + Cacheu1 + Cacheu3 + Bissau1); <i>culum</i> (CC + Cacheu2)	CVCVC	—	onom.	No	<i>fundá</i>	‘dive, sink’	v.	—
(9)	CC + 5	*/cut/	<i>cut</i> (CC + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	—	intens.	No	<i>melá</i>	‘be sweet’	v.	—
(10)	CC + 5	*/fa(r)kat/ ~ */fe(r)ket/	<i>fəkət</i> (CC); <i>fakat</i> (Cacheu1); <i>fərket</i> ~ <i>ferket</i> (Geba); <i>fargat</i> (Bissau1); <i>fat</i> (Cacheu2); <i>fat</i> (+ <i>kordá</i> ‘wake up’) (Bissau1); <i>firgit</i> (Cacheu3 + Bissau2)	CVCVC	Mandinka + Wolof©	immdt.	No	<i>labantá</i>	‘stand up’	v.	<i>caŋ</i> (Geba)
(11)	CC + 6	*/fandaŋ/	<i>fandaŋ</i> (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVCCVC	—	intens.	No	<i>braŋku</i>	‘be white’	qual.	—
(12)	CC + 6	*/fep/	<i>fep</i> (CC + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3); <i>fep</i> (Geba)	CVC	Mandinka + Wolof© (QM)	intens.	No	<i>kabá</i>	‘finish’	v.	—
(13)	CC + 7	*/fit/	<i>fit</i> (CC + Geba + Cacheu1 + Bissau1 + Bissau2 + Bissau3); <i>fis</i> (Cacheu2); <i>fif</i> (Cacheu3); <i>fas</i> (Bissau2); <i>falas</i> (Geba)	CVC	Mandinka©	immdt.	No	<i>pasá</i>	‘go by’	v.	—
(14)	CC + 5	*/fitfat/	<i>fitfat</i> (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1)	CVCCVC	Mandinka©	intens. + onom.	Yes	<i>jantí</i>	‘be hurried’	v.	—
(15)	CC + 6	*/fut/	<i>fut</i> (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau3); <i>fat</i>	CVC	Mandinka + Wolof©	immdt.	No	<i>sey</i>	‘get out’	v.	—

Continental UGC common ideophone											Other CUGC equiv.
No.	Distr. profile	Ideophone	Attested forms	SS	Etymon	SEM	RED	Form	Meaning	PoS	
			(Cacheu3 + Bissau1); <i>fup</i> (Bissau3); <i>fis</i> (Bissau3)								
(16)	CC + 6	*/jɔp/	<i>yɔp</i> (Geba + Cacheu1 + Cacheu2 + Bissau1); <i>yop</i> (CC + Cacheu3 + Bissau3)	CVC	Wolof©	intens.	No	<i>mojá</i>	‘get wet’	v.	<i>pɔcɔk</i> (Geba)
(17)	CC + 6	*/jem/	<i>yem</i> (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	Mandinka©	intens.	No	<i>friyá</i>	‘be cold’	qual.	—
(18)	CC + 6	*/kaŋ/	<i>kaŋ</i> (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1); <i>kaŋ</i> (+ <i>seku</i> ‘be dry’ (Bissau3); <i>kaŋkaraŋ</i> (Bissau1)	CVC	Wolof©	intens.	No	<i>risu</i>	‘be hard/ tough’	qual.	—
(19)	CC + 4	*/kat/	<i>kak</i> (CC); <i>kat</i> (Cacheu1 + Cacheu3); <i>pakat/ fakat</i> (Geba); <i>pargat</i> (Cacheu2); <i>fargat</i> (Cacheu3)	CVC	—	intens.	No	<i>kebrá</i>	‘break’	v.	—
(20)	CC + 6	*/keŋkereŋ/	<i>keŋkereŋ</i> (CC + Geba + Cacheu1 + Cacheu3 + Bissau1 + Bissau3); <i>kereŋ</i> (Cacheu2); <i>keŋ</i> (Geba)	CVCCVCVC	—	intens.	Yes	<i>saŋ</i>	‘be healthy’	qual.	—
(21)	CC + 6	*/kondoŋ/	<i>kondoŋ</i> (CC + Geba + Cacheu1 + Cacheu2 + Bissau3); <i>kindiŋ-kondoŋ</i> (Cacheu3 + Bissau1); <i>konkoroy</i> (Geba)	CVCCVC	—	atten. ‘completely alone’	No	<i>a-mi</i> , <i>a-bo...</i>	‘I, you...’	pron.	—
(22)	CC + 4	*/koroc/	<i>koroc</i> (CC + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1) <sup>48</sup>	CVCVC	—	atten.	No	<i>partí</i>	‘give’	v.	—
(23)	CC + 6	*/kuŋ/	<i>kuŋ</i> (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	—	intens.	No	<i>yĩncí</i>	‘be full’	v.	—

<sup>48</sup> A related form *koroc* ~ *ñoroc* was found in Geba (Geba) where it seems to be used only as an adjective (not as an ideophonic adverb).

Continental UGC common ideophone											Other CUGC equiv.
No.	Distr. profile	Ideophone	Attested forms	SS	Etymon	SEM	RED	Form	Meaning	PoS	
(24)	CC + 5	*/lot/	<i>lot</i> (CC + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1); <i>lt</i> (Geba)	CVC	—	intens.	No	<i>negá</i>	‘refuse’	v.	—
(25)	CC + 6	*/nok/	<i>nok</i> (CC + Cacheu3 + Bissau1 + Bissau3); <i>nɔk</i> (Geba + Cacheu1 + Cacheu2)	CVC	Mandinka© (QM)	intens.	No	<i>pretu</i>	‘be black’	qual.	—
(26)	CC + 4	*/nu(ŋ)/	<i>nu</i> (Geba + Cacheu1 + Cacheu3 + Bissau1); <i>nuŋ</i> (CC)	CVC	—	intens.	No	<i>kurpu-limpu</i>	‘be naked’	qual	—
(27)	CC + 4	*/palaw/ ~ */paraw/	<i>palaw</i> (CC); <i>falaw</i> (Geba); <i>paraw</i> (Cacheu3 + Bissau2); <i>paw</i> (Cacheu1)	CVCVC	—	visual	No	<i>lamprá</i>	‘shine’	v.	<i>pas ~ feles</i> (Geba)
(28)	CC + 4	*/parac/	<i>parac</i> (CC + Geba + Cacheu1 + Cacheu3); <i>parat</i> (Bissau1)	CVCVC	—	onom.	No	<i>kokó</i>	‘defecate’	v.	—
(29)	CC + 4	*/pat/ ~ */put/	<i>pat</i> (CC + Cacheu3 + Bissau1); <i>fat</i> (Cacheu1); <i>parat</i> (Bissau1); <i>put</i> (CC + Geba)	CVC	Mandinka + Wolof©	intens.	No	<i>sapá</i>	‘cut’	v.	—
(30)	CC + 6	*/pɔɔk/	<i>pɔɔk</i> (Cacheu1 + Cacheu2 + Cacheu3 + Bissau1); <i>potok</i> (Bissau3); <i>botok</i> (CC); <i>pɔɔk</i> (Geba)	CVCVC	Mandinka©	intens.	No	<i>moli</i>	‘be soft’	qual.	—
(31)	CC + 5	*/pɔɔk/ ~ */pucuk/	<i>potok</i> (CC); <i>pɔɔk</i> (Geba + Cacheu1); <i>putuk</i> (Cacheu2); <i>pɔɔk</i> (Geba); <i>pucuk</i> (Cacheu3 + Bissau1); <i>pucak</i> (Bissau3)	CVCVC	Mandinka©	intens.	No	<i>susu</i>	‘be dirty’	qual.	—
(32)	CC + 6	*/pus/	<i>pus</i> (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	—	intens.	No	<i>limpu</i>	‘be clean’	qual.	—
(33)	CC + 5	*/rap/	<i>rap</i> (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3); <i>rip</i> (Cacheu3); <i>raj</i> (Bissau1)	CVC	Wolof©	intens.	No	<i>ficá</i>	‘close’	v.	<i>varj</i> (Bissau1)

Continental UGC common ideophone										Associated element	Other CUGC equiv.
No.	Distr. profile	Ideophone	Attested forms	SS	Etymon	SEM	RED	Form	Meaning	PoS	
(34)	CC+6	*/sip/	<i>sip</i> (Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau2); <i>lip</i> (CC); <i>lip</i> (+ <i>fartá</i> ‘be satisfied’) (Cacheu3); <i>vip</i> (Geba)	CVC	—	intens.	No	<i>pisadu</i>	‘be heavy’	qual.	—
(35)	CC+4	*/taw/	<i>taw</i> (+ <i>malgós</i> ‘be bitter’) (CC); <i>taw</i> (+ <i>forti</i> ‘be sour’) (Geba + Cacheu1 + Cacheu2 + Cacheu3)	CVC	Mandinka + Wolof©	intens.	No	<i>malgós</i>	‘be bitter’	qual.	<i>rok</i> (+ <i>malgós</i> ‘be bitter’) (Geba)
(36)	CC+5	*/tip/	<i>tip</i> (CC + Geba + Cacheu1 + Cacheu3 + Bissau1); <i>tip</i> (+ <i>sukuru</i> ‘be dark’) (Bissau3); <i>sip</i> (Geba)	CVC	—	intens.	No	<i>pegá</i> (CC)/ <i>duru</i> (others)	‘thicken’	v.	—
(37)	CC+6	*/wak/	<i>wak</i> (Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3); <i>wɛk</i> (CC)	CVC	—	intens.	No	<i>burmeju</i>	‘be red’	qual.	—
(38)	CC+6	*/wandaŋ/	<i>wandaŋ</i> (CC + Cacheu3 + Bissau1 + Bissau2); <i>wantaŋ</i> (+ <i>klaru</i> ‘be clear(ed)’) (Geba); <i>wandalaŋ</i> (Cacheu1 + Cacheu2)	CVCCVC	Mandinka© (QM)	intens. (‘wide’)	No	<i>yabrí</i>	‘open (door)’	v.	—
(39)	CC+6	*/wit/	<i>wit</i> (CC + Geba + Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	Mandinka©	intens.	No	<i>kintí</i>	‘be hot’	qual.	—

**Table 1.4:** ‘other distribution’ type: ‘CC vs. Guinea-Bissau’ and ‘no CC vs. Guinea-Bissau’ subtypes.

No.	Distr. profile	Continental UGC ideophones						Associated element			Other CUGC equiv.	
		CC	Guinea-Bissau	Attested forms	SS	Etymon	SEM	RED	Form	Meaning		PoS
(1)	CC vs. Guinea-Bissau (3) <sup>49</sup>	[lɛm]	*/waŋ/	<i>waj</i> (Geba + Cacheu3 + Bissau1)	CVC	—	intens.	No	<i>didiya</i>	‘at noon’	adv.	—
(2)	CC vs. Guinea-Bissau (3)	[maj]	*/fis/	<i>fis</i> (Cacheu1 + Cacheu2); <i>fis</i> (+ <i>sai</i> ‘get out’) (Bissau3)	CVC	Mandinka ©	speech (‘without a word’)	No	<i>pasá</i>	‘go by’	v.	—
(3)	CC vs. Guinea-Bissau (4)	[liŋ]	*/kaŋ/	<i>kaŋ</i> (Geba + Cacheu2 + Bissau1); <i>taŋ</i> (Cacheu3)	CVC	Wolof©	intens.	No	<i>tesu</i>	‘be solid’	qual.	<i>ŋaw</i> (Geba)
(4)	no CC vs. Guinea-Bissau (2)	—	*/sip/	<i>sip</i> (Bissau3); <i>zip</i> (Geba)	CVC	—	intens.	No	<i>gros</i>	‘be fat/thick’	qual.	—
(5)	no CC vs. Guinea-Bissau (4)	—	*/far/	<i>far</i> (Geba + Cacheu1 + Cacheu2 + Bissau3)	CVC	Wolof©	intens.	No	<i>negá</i>	‘refuse’	v.	—
(6)	no CC vs. Guinea-Bissau (5)	—	*/caw/ <sup>50</sup>	<i>caw</i> (Geba + Cacheu2 + Bissau3); <i>cadaw</i> (Cacheu3 + Bissau1 + Bissau3)	CVC	Mandinka + Wolof©	intens. (‘ripe fruit’)	No	<i>burmeju</i>	‘red’	qual.	—

<sup>49</sup> In Tables 1.4 and 1.5, the digits given in brackets in the column ‘Distribution profile’ correspond to the number of consultants who supported a given form.

<sup>50</sup> The form \*/caw/ has been retained as the base form for this ideophone against *cadaw*, which is almost certainly an extended (expressive) form of *caw* (see also \*/baw/ above).

Table 1.5: ‘other distribution’ type: ‘(CC + Guinea-Bissau) vs. Guinea-Bissau’ subtype.

Continental UGC ideophones														Other CUGC equiv.		
														Associated element		
														Form	Meaning	PoS
(CC + Guinea-Bissau)				Guinea-Bissau		SS	Etymon	SEM	RED							
No.	Distr. profile	CF	Attested forms	CF	Attested forms											
(1)	(CC + Guinea-Bissau) (2) vs. Guinea-Bissau (3)	*/mut/	<i>mut</i> (CC+Geba)	*/mip/	<i>mip</i> (Cacheu3 + Bissau2 + Bissau3)	CVC	Wolof©	intens.	No	<i>sukuru</i>	‘be dark’	qual.	—			
(2)	(CC + Guinea-Bissau) (3) vs. Guinea-Bissau (3)	*/ɲos/	<i>ños</i> (CC+ Cacheu3); <i>ñɔk</i> (Geba)	*/rek/	<i>rek</i> (Cacheu3 + Bissau1); <i>rek</i> (+ <i>justá</i> ‘be enough’) (Bissau3)	CVC	Wolof©	intens.	No	<i>cigá</i>	‘be closely related (parents)’	v.	—			
(3)	(CC + Guinea-Bissau) (2) vs. Guinea-Bissau (5)	*/mik/	<i>mik</i> (CC); <i>muruk</i> (Geba)	*/jem/	<i>yem</i> (Cacheu1 + Cacheu2 + Cacheu3 + Bissau1 + Bissau3)	CVC	Mandinka + Wolof©	intens.	No	<i>kalá</i>	‘keep quiet’	v.	<i>tik</i> (Cacheu2)			



Appendix 2: Insular UGC ideophones

Table 2.1: Santiago ideophones.

Santiago ideophone					Associated element				Continental UGC ideophone		
No.	Ideophone	SS	Etymon	Semantic type	RED	Form	Meaning	PoS	Semantic Equiv. <sup>51</sup>	Cognate	Distr. type
(1)	<i>álbu</i> [ˈalbu]	VCCV	Port. (§4.4.1)	intens.	No	<i>bráunku</i>	‘white’	adj.	*/fandaŋ/	No	common core
(2)	<i>ban</i> [ˈbɛ̃]	CVC	—	intens.	No	<i>intxi</i>	‘fill’ ~ ‘be full’	v.	*/kuŋ/	No	common core
(3)	<i>bipu</i> [ˈbipu] ~ <i>vipu</i> [ˈvipu]	CVCV	—	intens.	No	<i>intxi</i>	‘fill’ ~ ‘be full’	v.	*/kuŋ/	No <sup>52</sup>	common core
(4)	<i>bódji-bódji</i> [ˈbɔdʒi ˈbɔdʒi]	CVCV-CVCV	Mandinka©	intens.	Yes	<i>gordu</i>	‘fat’	adj.	*/sip/	No	other
(5)	<i>bupu</i> [ˈbupu]	CVCV	—	onom.	No	<i>kai</i>	‘fall’	v.	*/bup/	Yes	common core
(6)	<i>fakati</i> [feˈketi]	CVCVCV	Mandinka + Wolof©	immdt.	No	<i>labánta</i>	‘stand up’	v.	*/ <b>fa</b> (r)kat/ ~ */ <b>fe</b> (r)ket/	Yes	common core
(7)	<i>fépu</i> [ˈfɛpu]	CVCV	Mandinka + Wolof© (QM)	intens.	No	<i>kumi/kába</i>	‘eat, finish’	v.	*/fep/	Yes	common core

<sup>51</sup> In the column ‘semantic equivalent’, the segments printed in bold justify the existence of an etymological relationship between insular and continental UGC ideophones, and they have lead us to write ‘Yes’ in the column ‘cognate’.

<sup>52</sup> A possible link could be posited with the continental form \*/sip/, which intensifies *pisadu* ‘be heavy’ (see Table 1.3), a lexical item semantically close to ‘be full’. Furthermore, we have noted a variant *vip* [vip] in Geba, which displays an even greater similarity with Capeverdean *vipu*. However, as *vip* was only attested once and as the semantic relationship between both associated elements (*pisadu* and *intxi*) is not absolutely straightforward, we have preferred not to consider these two ideophones as cognates within the scope of this study.

Santiago ideophone					Associated element				Continental UGC ideophone		
No.	Ideophone	SS	Etymon	Semantic type	RED	Form	Meaning	PoS	Semantic Equiv. <sup>51</sup>	Cognate	Distr. type
(8)	<i>finu</i> [ˈfinu]	CVCV	Mandinka + Port.© (QM)	intens.	No	<i>prétu</i>	‘black’	adj.	*/nok/	No	common core
(9)	<i>futi</i> [ˈfuti]	CVCV	Mandinka + Wolof©	immdt.	No	<i>sai</i>	‘get out’	v.	*/fut/	Yes	common core
(10)	<i>fututú</i> [futuˈtu]	CVCVCV	—	intens.	Yes	<i>txera</i>	‘smell’ (intr.)	v.	[fututut]	Yes	CC-only
(11)	<i>góbu-góbu</i> [ˈgɔbu ˈgɔbu]	CVCV-CVCV	Mandinka© (QM)	intens.	Yes	<i>bedju</i>	‘old’	adj.	*/kokorot/	No	minority
(12)	<i>ien</i> [ˈjɛ] ~ <i>ian</i> [ˈjɛ]	CVC	Mandinka + Wolof©	intens.	No	<i>kála</i>	‘be silent’	v.	*/jem/	Yes	other
(13)	<i>iondon</i> [jõˈdõ]	CVCCVC	Wolof©	visual (‘desert’)	No	<i>fika</i>	‘be’	v.	*/wandalan/	Yes	minority
(14)	<i>kati</i> [ˈketi]	CVCV	—	intens.	No	<i>kebra</i>	‘break’	v.	*/kat/	Yes	common core
(15)	<i>kóti-kóti</i> [ˈkɔti ˈkɔti]	CVCV-CVCV	Mandinka© (QM)	intens.	Yes	<i>bedju</i>	‘old’	adj.	*/kokorot/	No	minority
(16)	<i>lápu</i> [ˈlapu]	CVCV	Wolof©	intens. (‘tight’)	No	<i>xinta</i>	‘sit’	v.	*/bap/ ~ [vap] (Bissau1)	Yes	minority
(17)	<i>mápu</i> [ˈmapu]	CVCV	—	intens.	No	<i>pánha / pega / da</i>	‘seize, take’	v.	[mɛp]	Yes	CC-only
(18)	<i>moku</i> [ˈmoku]	CVCV	Wolof© (QM)	intens.	No	<i>pila/kebra</i>	‘crush’	v.	*/moku/ (Bissau3 + CC)	Yes	NA
(19)	<i>muku-muku</i> [ˈmuku ˈmuku]	CVCV-CVCV	Mandinka + Wolof©	onom. (‘silent’)	Yes	<i>fika</i>	‘be’	v.	[mik] ~ [muruk]	Maybe	other
(20)	<i>mus-mus</i> [ˈmus ˈmus]	CVC-CVC	Wolof©	intens.	Yes	<i>perdi</i>	‘disappear’	v.	[mes]	Yes	CC-only

Santiago ideophone					Associated element				Continental UGC ideophone		
No.	Ideophone	SS	Etymon	Semantic type	RED	Form	Meaning	PoS	Semantic Equiv. <sup>51</sup>	Cognate	Distr. type
(21)	<i>(nun)priti</i> [nũ'priti]	CCVCV	—	intens. (‘stark naked’)	No	<i>fika</i>	‘be’	v.	*/nu(ŋ)/	No	common core
(22)	<i>pan-pan</i> [pẽ 'pẽ]	CVC-CVC	—	intens.	Yes	<i>ségu</i>	‘blind’	adj.	—	Zero	NA
(23)	<i>plátxi-plátxi</i> [ˈplatʃi ˈplatʃi]	CCVCV- CCVCV	—	intens.	Yes	<i>gordu</i>	‘fat’	adj.	*/sip/	No	other
(24)	<i>pran-pran</i> [ˈprẽ ˈprẽ]	CCV-CCV	—	speech (‘clearly’)	Yes	<i>fla/papia</i>	‘say, talk’	v.	[pɛjɛj]	No	CC-only
(25)	<i>pratxi</i> [ˈpretʃi]	CCVCV	—	onom.	No	<i>fasi kokó</i>	‘defecate’	VP	*/parac/	Yes	common core
(26)	<i>prikiti</i> [priˈkiti]	CCVCVCV	Mandinka + Wolof©	immdt.	No	<i>labánta</i>	‘stand up’	v.	*/pirkit/	Yes	minority
(27)	<i>pufu</i> [ˈpufu]	CVCV	—	intens.	No	<i>fra</i>	‘make a hole, pierce’	v.	*/puf/	Yes	minority
(28)	<i>saki</i> [ˈseki] ~ <i>siki</i> [ˈsiki]	CVCV	—	intens.	No	<i>sápa</i>	‘get cut’	v.	*/pat/ ~ */put/	No	common core
(29)	<i>tran</i> [ˈtrẽ]	CCVC	—	intens.	No	<i>sakédu</i>	‘stand’	v.	*/caŋ/ ~ */kaŋ/ ~ [taŋ] (Bissau1)	Maybe	other
(30)	<i>txápu</i> [ˈtʃapu]	CVCV	Mandinka©	intens.	No	<i>pánha / pega / da</i>	‘seize, take’	v.	*/cap/	Yes	common core
(31)	<i>txulupu</i> [tʃuˈlupu] ~ <i>txulufu</i> [tʃuˈlufu] ~ <i>txufun</i> [tʃuˈfũ]	CVCVCV ~ CVCVCV ~ CVCVC	—	onom. (‘splash’)	No	<i>da</i>	‘give’	v.	*/culup/	Yes	common core

Santiago ideophone					Associated element				Continental UGC ideophone		
No.	Ideophone	SS	Etymon	Semantic type	RED	Form	Meaning	PoS	Semantic Equiv. <sup>51</sup>	Cognate	Distr. type
(32)	<i>uandan</i> [wě'dě] ~ <i>iandan</i> [jě'dě]	CVCCVC	Mandinka© (QM)	intens.	No	<i>abri</i>	'open'	v.	*/wandaŋ/	Yes	common core
(33)	<i>uátchi</i> ['watʃi]	CVCV	Mandinka©	intens.	No	<i>fendi</i>	'split open'	v.	*/wac/	Yes	minority
(34)	<i>uís</i> ['wis]	CVC	—	intens.	No	<i>lolu</i>	'slip'	v.	[fit] (CC)	No	NA

## Appendix 3: Etymology of some UGC ideophones

Table 3.1: Origins of those African-derived ideophones for which a possible etymon was identified.

UGC	Proto-CUPC form */ or Santiago [] form	Origin	Table	Distribution category	Etymon
Continental	*/caw/	Mandinka + Wolof	1.4	other (no CC vs. Guinea-Bissau)	Mandinka <i>cáw</i> (IDEO) ‘very [red], intensely [red]’ (Creissels 2012: 33); Wolof <i>coy</i> [cɔj] (IDEO) ‘very [red]’ (Diouf 2003: 85)
Continental	[colop]	Mandinka	1.1	CC-only	Mandinka <i>cúrúm</i> (IDEO) ‘[jump] quickly [down]’ (Creissels 2012: 34)
Continental	*/corot/	Mandinka	1.3	common core	Mandinka <i>córóti</i> (v.) ‘drip, be tiny’ (Creissels 2012: 34)
Continental	*/far/	Wolof	1.4	other (no CC vs. Guinea-Bissau)	Wolof <i>fétt</i> [fɛ:ttə] (IDEO) ‘[refuse] adamantly, strongly [refuse]’ (Diouf 2003: 126)
Continental	*/fas/	Wolof	1.2	minority	Wolof <i>fés</i> [fes] (v.) ‘be visible/conspicuous’ (Diouf 2003: 126)
Continental	*/fis/	Mandinka	1.4	other (CC vs. Guinea-Bissau)	Mandinka <i>fíw</i> (IDEO) ‘[go by] quickly’ (Creissels 2012: 70)
Continental	*/fit/	Mandinka	1.3	common core	Mandinka <i>fit</i> (IDEO) ‘[go by] quickly’ (Creissels 2012: 69); <i>fálás</i> (IDEO) ‘[go by] quickly’ (Creissels 2012: 58)
Continental	*/fitfat/	Mandinka	1.3	common core	Mandinka (IDEO) <i>páfát</i> ~ <i>pápát</i> ‘[flee] quickly’ (Creissels 2012: 200)
Continental	[fof]	Mandinka	1.1	CC-only	Mandinka <i>yéf</i> (IDEO) ‘very [light]’ (Creissels 2012: 271)
Continental	*/fuf/ ~ */fus/	Wolof	1.2	minority	Wolof <i>fuuf</i> [fu:f] (adv.) ‘much [more]’ (Diouf 2003: 132)
Continental	*/jem/	Mandinka	1.3	common core	Mandinka <i>yóm</i> [jóm] (IDEO) ‘very [cold]’ (Creissels 2012: 273)
Continental	*/jɔp/	Wolof	1.3	common core	Wolof <i>xəpp</i> [xəppə] (IDEO) ‘completely [wet]’ (Diouf 2003: 387; Fal et al. 1990: 256)
Continental	[kamaj]	Wolof	1.1	CC-only	Wolof <i>kamaj</i> [kamaɕ] (IDEO) ‘[be] suddenly [switched off]’ (Diouf 2003: 181)
Continental	*/kaŋ/	Wolof	1.3 + 1.4	common core + other (CC vs. Guinea-Bissau)	Wolof <i>kěj</i> [kəŋ] (IDEO) ‘very [hard/tough]’ (Diouf 2003: 185)

UGC	Proto-CUPC form */ or Santiago [] form	Origin	Table	Distribution category	Etymon
Continental	[koŋ]	Wolof	1.1	CC-only	Wolof <i>koŋ</i> [kɔŋ] (IDEO) ‘very [dry]’ (Diouf 2003: 188)
Continental	*/las/	Wolof	1.2	minority	Wolof <i>lacc</i> [la:ccə] (v.) ‘be exhausted’ (Diouf 2003: 193)
Continental	*/mut/	Wolof	1.5	other ((CC + Guinea-Bissau) vs. Guinea-Bissau)	Wolof <i>muut</i> [mu:t] (v.) ‘keep calm, stand still’ (Diouf 2003: 231)
Continental	*/nok/	Mandinka	1.3	common core	Mandinka <i>nót</i> (IDEO) ‘very (black and dirty)’ (Creissels 2012: 188)
Continental	*/pat/ ~ */put/	Mandinka + Wolof	1.3	common core	Mandinka <i>pát</i> (IDEO) ‘[cut] quickly’ (Creissels 2012: 201); <i>pókót</i> ‘[cut] silently’ (Creissels 2012: 203); Wolof <i>pacc</i> (IDEO) ‘[cut] into two equal parts’ (Fal et al. 1990: 256)
Continental	*/pɔ̀tək/	Mandinka	1.3	common core	Mandinka <i>páták</i> (IDEO) ‘very [wet]’ (Creissels 2012: 201); <i>pòtó</i> (n.) ‘clay, mud, silt’ (Creissels 2012: 203)
Continental	*/pɔ̀tək/ ~ */pucuk/	Mandinka	1.3	common core	Mandinka <i>hót</i> (IDEO) ‘very [black and dirty]’ (Creissels 2012: 82)
Continental	[pul]	Wolof	1.1	CC-only	Wolof <i>pëll</i> [pəllə] (IDEO) ‘[get out] quickly’ (Diouf 2003: 277-278; Fal et al. 1990: 169)
Continental	*/rap/	Wolof	1.3	common core	Wolof <i>rapp</i> [ra:ppə] (IDEO) ‘[close] tightly’ (Diouf 2003: 285)
Continental	*/rek/	Wolof	1.5	other ((CC + Guinea-Bissau) vs. Guinea-Bissau)	Wolof <i>rekk</i> [re:kkə] (adv.) ‘only’ (Diouf 2003: 289); <i>rəkk</i> [rəkkə] ~ <i>rikk</i> [rikkə] (IDEO) ‘just [beside], right [next to]’ (Diouf 2003: 289, 291)
Continental	*/taw/	Mandinka + Wolof	1.3	common core	Mandinka <i>láv</i> (IDEO) ‘very [bitter]’ (Creissels 2012: 156); Wolof <i>toll</i> [tɔ̀llə] (IDEO) ‘very [bitter]’ (Diouf 2003: 347)
Continental	*/tep/	Mandinka	1.2	minority	Mandinka <i>tép</i> (IDEO) ‘completely [full], [filled] to the brim’ (Diouf 2003: 58)
Continental	[was]	Mandinka + Wolof	1.1	CC-only	Mandinka <i>wásák</i> (IDEO) ‘[pour/spill] all over’, expresses the idea of ‘dispersal’ or ‘scattering’ (Creissels 2012: 264)

UGC	Proto-CUPC form */* or Santiago [] form	Origin	Table	Distribution category	Etymon
					and/or Wolof <i>wesar</i> [wɛsar] (v.) ‘be scattered’ (Diouf 2003: 368)
Continental	*/wit/	Mandinka	1.3	common core	Mandinka <i>wij</i> [wij] (IDEO) ‘very [hot]’ (Creissels 2012: 265)
Continental + Insular	*/fep/ + [ʼfɛpu]	Mandinka + Wolof	1.3 + 2.1	common core	Mandinka <i>fɛw</i> (IDEO) ‘[finish] completely’ (Creissels et al. 1982: 47, Rougé 2004: 308) and/or Wolof <i>-épp</i> (adv.) ‘completely’; <i>fépp</i> [fɛ:ppə] (adv.) ‘complete, every(where)’ (Diouf 2003: 125)
Continental + Insular	*/wandaŋ/ + [wɛ̃ˈdɛ̃] ~ [jɛ̃ˈdɛ̃]	Mandinka	1.3 + 2.1	common core	Mandinka <i>wáráj</i> (IDEO) ‘[swing] wide open’ (Creissels 2012: 264)
Continental + Insular	*/moku/ + [ˈmoku]	Wolof	2.1	NA	Wolof <i>mokk</i> [mɔːkkə] (v.) ‘be ground/crushed’ (Dieng 1985: 247; Diouf 2018: 226)
Continental + Insular	*/fut/ + [ˈfuti]	Mandinka + Wolof	1.3 + 2.1	common core	Mandinka <i>farafat</i> (IDEO) ‘[come out] suddenly’ (Creissels 2012: 61) and/or Wolof <i>poset</i> (IDEO) ‘(come out) suddenly’ (Fal et al. 1990: 171)
Continental + Insular	*/jem/ + [ˈjɛ̃] ~ [ˈjɛ̃]	Mandinka + Wolof	1.5 + 2.1	other ((CC + Guinea-Bissau) vs. Guinea-Bissau)	Mandinka <i>yéréŋ</i> (IDEO) ‘very [quiet]’ and/or Wolof <i>remm</i> [rɛːmmə] (IDEO) ‘quiet(ly) (sea)’ (Diouf 2003: 289); <i>xerem</i> [xɛrɛm] ‘completely [quiet]’ (Diouf 2003: 387)
Continental + Insular	*/yambalaŋ/ (CC) ~ */wandalan/ (Guinea-Bissau) + [jɔ̃ˈdɔ̃] (Santiago) + [ɛ̃ˈdɛ̃] (Fogo)	Wolof	1.2 + 2.1	minority	Wolof <i>yambalaŋ</i> [jaːmbalaŋ] (IDEO) ‘wide [open], extensive, large’ (Diouf 2003: 398). The CC is obviously derived from Wolof. The other forms may have been influenced by Mandinka <i>wáráj</i> ‘wide [open]’ which is the most plausible etymon of continental */wandaŋ/ and insular [wɛ̃ˈdɛ̃] ~ [jɛ̃ˈdɛ̃] ‘wide (open)’ (Quint & Moreira Tavares 2019: 158). Some Santiago speakers consider that [wɛ̃ˈdɛ̃] ~ [jɛ̃ˈdɛ̃] can be used indifferently in lieu of [jɔ̃ˈdɔ̃] while others maintain a distinction [jɔ̃ˈdɔ̃] to refer to a ‘quiet, empty place’ vs. [wɛ̃ˈdɛ̃] to intensify the ‘opening of a door’. We have considered here that the two roots are separated, which is probably historically true.

UGC	Proto-CUPC form */* or Santiago [] form	Origin	Table	Distribution category	Etymon
Continental + Insular	*/bap/ + ['lapu]	Wolof	1.2 + 2.1	minority	Wolof <i>mbàpp</i> [mba:ppə] (IDEO) (Diouf 2003: 217) ~ <i>wàpp</i> (IDEO) (Fal et al. 1990: 239) 'sit down heavily'
Continental + Insular	*/mik/ + ['muku 'muku]	Mandinka + Wolof	1.2 + 2.1	other ((CC + Guinea-Bissau) vs. Guinea-Bissau)	Mandinka <i>měj</i> (IDEO) 'very [quiet]' (Creissels 2012: 170) and/or Wolof <i>müig</i> [mi:k] (itj.) 'hush, shh' (Diouf 2003: 225); <i>tëkk</i> [tɛ:kkə] 'very (quiet)'
Continental + Insular	[mes] + ['mus 'mus]	Wolof	1.1 + 2.1	CC-only	Wolof <i>mes</i> [mɛs] (IDEO) '[disappear] suddenly' (Diouf 2003: 225; Fal et al. 1990: 131)
Continental + Insular	*/pirkit/ + [pri'kiti]	Mandinka	1.2 + 2.1	minority	Mandinka <i>kíríbít</i> (IDEO) '[stand up] suddenly' (Creissels 2012: 127)
Continental + Insular	*/cap/ + ['tʃapu]	Mandinka	1.3 + 2.1	common core	Mandinka <i>cápi</i> (v.) 'take, seize' (Creissels 2012: 33); <i>cás</i> (IDEO) '[catch] brutally [in the air]' (Creissels 2012: 33); <i>táp</i> (IDEO) '[seize] quickly' (Creissels 2012: 244)
Continental + Insular	*/wac/ + ['watʃi]	Mandinka	1.2 + 2.1	minority	Mandinka <i>wác</i> ~ <i>wáj</i> (IDEO) 'rip, brutally [torn]' (Creissels 2012: 263)
Continental + Insular	*/fa(r)kat/ ~ */fe(r)ket/ + [fe'keti]	Mandinka + Wolof	1.3	common core	Mandinka <i>cókót</i> (IDEO) '[stand up] at once' (Creissels 2012: 33); <i>táráfát</i> (IDEO) '[stand up] suddenly' (Creissels 2012: 245) and/or Wolof <i>ñokket</i> [ɲo:kket] (IDEO) '[stand up] at once' (Diouf 2003: 269); <i>fojjet</i> [fɔ:ʝet] (IDEO) '[stand up] immediately' (Diouf 2003: 128)
Insular	['bɔdʒi 'bɔdʒi]	Mandinka	2.1	common core	Mandinka <i>bòjboji</i> (v.) 'be too ripe [fruit], too oily [dish]' (Creissels et al. 1982: 16)
Insular	['finu]	Mandinka + Port.	2.1	common core	Mandinka <i>fñj</i> (IDEO) 'black' (Creissels 2012: 67; Rougé 2004: 309) and/or Port. <i>fino</i> ['finu] (adj.) 'fine, refined'
Insular	['gɔbu 'gɔbu]	Mandinka	2.1	minority	Mandinka <i>kób</i> (IDEO) 'very (old)' (Creissels 2012: 128)
Insular	['kɔti 'kɔti]	Mandinka	2.1	minority	Mandinka <i>kòtó/kòtóo</i> (qual./n.) '(be) old, elder' (Creissels 2012: 137)



**Abbreviations (only used in the Appendices)**

© = African etymon given in Appendix 3	immdt. = immediate	(QM) = published in Quint &
adj. = adjective	itj. = interjection	Moreira Tavares (2019: 157-158)
atten. = attenuation	intens. = intensifying/intensive	qual. = qualifier
CF = common form	n. = noun	SEM = semantic type
CUGC = Continental Upper Guinea Creole	onom. = onomatopoeia (= sound ideophone)	SS = syllable structure
Distr. = distribution	PoS = Part of Speech	v. = verb
equiv. = equivalent	pron. = pronoun	VP = verb phrase