Universiteit
Leiden
The Netherlands

A lexical and phonological reconstruction of Highland East Cushitic including a comparative East Cushitic database

Sosal Altayeb Mohammed Ali, A.

Citation
Sosal Altayeb Mohammed Ali, A. (2026, July 9). *A lexical and phonological reconstruction of Highland East Cushitic including a comparative East Cushitic database*. *LOT dissertation series*. LOT, Amsterdam. Retrieved from https://hdl.handle.net/1887/4307764

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4307764

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

A Lexical and Phonological Reconstruction of Highland East Cushitic

Published by
LOT
Binnengasthuisstraat 9
1012 ZA Amsterdam
The Netherlands

phone: +31 20 525 2461

e-mail: lot@uva.nl
http://www.lotschool.nl

Cover illustration: Cover photo taken by Rahel T. Dires on 11 October 2023 in Durame, Kambaata Region, Ethiopia.

ISBN: 978-94-6093-506-0
DOI: https://dx.medra.org/10.48273/LOT0721
NUR: 616

Copyright © 2026: Ahmed Sosal. All rights reserved.

A Lexical and Phonological Reconstruction of Highland East Cushitic

including a Comparative East Cushitic Database

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus, Prof.dr. S. de Rijcke,
volgens besluit van het college voor promoties
te verdedigen op donderdag 9 juli 2026
klokke 13:00 uur

door

Ahmed Sosal

geboren te Sennar, Soedan
in 1988

Promotor:	Prof.dr. Maarten Mous
Co-promotor:	Dr. Christian J. Rapold
Promotiecommissie:	Prof.dr. Maarten Kossmann
	Prof.dr. Mauro Tosco (University of Turin)
	Dr. Yvonne Treis (LLACAN, CNRS)
	Dr. Marijn van Putten

The research reported here was conducted as part of the Open Competition SSH project ‘Unravelling East Africa’s Early Linguistic History’ (LHEAf; 40618TW013), funded by the Netherlands Organization for Scientific Research (NWO).

Table of Contents

Table of Contents i
Tables iv
Figures iv
Abbreviations and Symbols v
Acknowledgements ix
1 Introduction 1
1.1 Background 1
1.1.1 Number of Speakers 1
1.1.2 Geographical Location 3
1.1.3 Classification 4
1.2 Literature and Data Sources 5
1.2.1 Data Compilation and Sources 6
1.2.2 Previous Comparative Work 7
1.3 East Cushitic Database 8
1.3.1 Overview and Scope 9
1.3.2 Organization and Presentation 10
1.3.3 Conventions and Visual Cues 12
1.3.4 Navigating the Database 12
1.3.5 Sources and Data Evaluation 13
1.3.6 Working Hypotheses and Data Interpretation 14
1.4 Research Rationale and Objectives 15
1.5 Methodological Framework 17
1.5.1 The Probability Scale for Reconstruction 17
1.5.2 Terminology and Analytical Framework 19
1.5.3 Data Presentation and Referencing Conventions 21
1.5.4 Contact and Inheritance as a Continuum 24
1.6 Structure of the Dissertation 25
1.7 Summary 28
2 Phonology 31
2.1 HEC Consonants 32
2.1.1 Pulmonic Alveolars and Post-alveolar 33

2.1.2 Non-pulmonic Alveolars and Post-alveolar 49
2.1.3 Labials 77
2.1.4 Glides 83
2.1.5 Velars, Glottals, and the Pharyngeal 89
2.2 HEC Vowels 113
2.2.1 Regular Stem Vowels 114
2.2.2 Variation in Stem Vowels 121
2.2.3 Terminal Vowels 131
2.2.4 Summary 140
2.3 Syllable and Root Structure 141
2.3.1 Syllable Onsets 141
2.3.2 Syllable Nuclei 142
2.3.3 Syllable Codas 143
2.3.4 Common Word and Root Structures 145
2.3.5 Summary 146
2.4 Accent 147
2.5 Summary and Conclusions 152
2.5.1 Consonant System 153
2.5.2 Reflexes of Reconstructed Consonants 155
2.5.3 Vowel System 158
2.5.4 Phonotactics and Prosody 159
3 Alternations at Morpheme Boundaries 161
3.1 Pulmonic Alveolars 162
3.1.1 Assimilation 162
3.1.2 Palatalization 174
3.2 Velars 183
3.2.1 Palatalization 183
3.2.2 Lenition 184
3.3 Glottal Stop 188
3.3.1 Fusion 189
3.3.2 Metathesis 203
3.3.3 Elision 206
3.3.4 Epenthesis 209

3.4 Sonorants 223
3.4.1 Epenthetic -i-, -y- 223
3.4.2 Epenthetic -n- 227
3.4.3 Epenthetic -l- 229
3.5 Summary and Conclusions 234
4 Morphology 239
4.1 Noun Morphology 239
4.1.1 Number 240
4.1.2 Deverbative 245
4.1.3 Problematic Extension 248
4.1.4 Genitive 250
4.1.5 Gender and Case 272
4.1.6 Personal Pronouns 285
4.2 Verb Morphology 293
4.2.1 The Passive 293
4.2.2 The Causative 294
4.2.3 The Reflexive 297
4.2.4 Support Verbs 307
4.2.5 Denominative Derivation 315
4.3 Summary and Conclusions 343
5 Conclusions 351
5.1 What the Study Has Established 351
5.1.1 Phonological Reconstruction 351
5.1.2 Morphological Reconstruction 352
5.2 Concluding Phonology Synthesis 352
5.2.1 Phonological System 353
5.2.2 Segmental Profile of PHEC 353
5.2.3 Phonotactics and Prosody 354
5.2.4 Boundary Alternations 354
5.2.5 HEC Sound Classes 355
5.3 Subgrouping Evidence 356
5.3.1 Phonological Diagnostics 358
5.3.2 Morphological Diagnostics 360

5.3.3 Subgrouping and Revised EC Labels 363
5.4 Language Contact and Diffusion 366
5.5 Methodological Implications 368
5.6 Contributions and Recommendations 370
5.7 Conclusion 371
References 373
Appendix 379

Tables

Table 1: Highland East Cushitic Consonant Inventory 33
Table 2: Highland East Cushitic Vowels Inventory 113
Table 3: Proto-Highland East Cushitic Consonants 153
Table 4: Reconstructed HEC Consonants and their Correspondences 155
Table 5: Proto-Highland East Cushitic Vowels 158
Table 6: Summary of Morphophonological Processes 235
Table 7: Reconstructed HEC Morphemes and Their Reflexes 345
Table 8: Reconstructed HEC Pronouns: Nominative, Accusative, and Possessive 348
Table 9: Major Sound Classes and Boundary Processes 355
Table 10: PHEC Consonant Inventory Organized by Class (place × manner) 356
Table 11: Morphophonological Subgrouping Diagnostics 359
Table 12: Morphological Subgrouping Diagnostics 362

Figures

Figure 1: Map of the Highland East Cushitic Languages 3
Figure 2: Highland East Cushitic Family Tree 4
Figure 3: Historical Development of the Genitive 271
Figure 4: The Development of the HEC Verbal Derivation System 343
Figure 5: Hudson's Family Vine Tree (2007, 1989, 1981) 363
Figure 6: Revised East Cushitic Subgroups Labels 365

Abbreviations and Symbols

Linguistic Glosses

Abbreviation	Meaning
1pl	first-person plural
1sg	first-person singular
2pl	second-person plural
2sg	second-person singular
3pl	third-person plural
3sg	third-person singular
acc	accusative
adj	adjective
caus1	first causative in double causative formation
caus2	second causative in double causative formation
coll	collective
dem	demonstrative
det	determiner
f	feminine
fn.	footnote, endnote
gen	genitive
interj	interjection
intr	intransitive
m	masculine
nom	nominative
obj	object
perf	perfective
pl	plural
plv	plurative
poss	possessive
pred	predicative
pron	pronoun
refl	reflexive
sg	singular
sgv	singulative
sp.	species
subj	subject
tr	transitive
TV	terminal vowel
voc	vocative

Languages and Language Groups

Abbreviation	Language/Group
Arb	Arbore
Bay	Bayso
CEC	Central East Cushitic
Dah	Dahalo
Das	Dhaasanac
Dir	Diraytata
EC	East Cushitic
Elm	Elmolo
EOT	Eastern Omo-Tana
Gaw	Gawwada
Ged	Gedeo
Had	Hadiyya
HEC	Highland East Cushitic
Kam	Kambaata
Kon	Konso
Mos	Mositacha
NEC	Northern East Cushitic
NHEC	Northern Highland East Cushitic
Oro	Oromo
OT	Omo-Tana
P	Proto-
PC	Proto-Cushitic
Ren	Rendille
SEC	Southern East Cushitic
Sid	Sidaama
Som	Somali
Ts’am	Ts’amakko
WOT	Western Omo-Tana
WR	West Rift

Symbols and Conventions

Symbol	Meaning
*X	reconstructed proto-form
- and +	morpheme boundary
-X-	inserted segment
=	clitic boundary
→ and ←	becomes / comes from
C, V	consonant, vowel in structural descriptions
Ch.	chapter
lit.	literary translation
X : Y	X corresponds to Y
X̥	devoiced segment
$X_{\overset{\circ}{i}}$	stem/root final segment
$\mathbf{V_{TV}}$	terminal vowel
~	reduplication (no spaces) / free variation (space before and after)
§	section (§§ two or more sections)

Acknowledgements

This work grew out of many conversations, corrections, and kindnesses; it has more authors than the cover suggests. What follows is the short version.

I am deeply grateful to my promoters, Christian Rapold and Maarten Mous. Christian started with me at the basics as a newcomer to comparative historical linguistics and stayed close through the slow weeks and the stuck arguments, listening and advising even after he moved away from Leiden. Special thanks to his sister, Noemi, who always came by to say “Hi” during our online meetings. Maarten’s project leadership, extensive knowledge of African languages, and unfailing kindness have shaped this project from start to finish. I also extend my thanks to his wife, Coos, who always opened her home to all of us.

The reading committee offered crucial comments that improved the manuscript. I highly appreciate your constructive feedback. I am especially indebted to Yvonne Treis, who first introduced me to Highland East Cushitic (HEC), invited and supported my Labex stay in Paris, generously shared Kambaata data, and gave sustained feedback on drafts. Your support and mentorship were indispensable.

The Early Linguistic History of East Africa (LHEAf) team has been a constant source of expertise and support. Thank you, Maarten Mous, Christian Rapold, Andrew Harvey, Samuel Beer, Terrill Schrock, Sjef van Lier, Dominique Loviscach, Giulia Zhang, Alba Hermida Rodríguez, Richard Griscom, Hope Morgan, Marta Cestari, Bettina Mourik, and others, for the long discussions and helpful suggestions. Your generosity in sharing knowledge and time made this project possible. I am particularly grateful for the friendship of Nina van der Vlugt and Jeroen van Ravenhorst (my paranymphs), and Iris Kruijsdijk.

Faculty and staff at Leiden University made life and work easier: Sara Petrollino (thanks for being helpful and kind), Maarten Kossmann, Felix Ameka, Azeb Amha, Nancy Kula, Jenneke van der Wal, Victoria Nyst, Hilde Gunnink, and Thilo Schadeberg; Jurgen Lingen, Maarit van Gammeren, Tijmen Pronk, Kate Bellamy, Michaël Peyrot, Sasha Lubotsky, Benjamin Suchard, Khalid Mourigh, and others. Special thanks to Heleen Smits for welcoming me when I first arrived in the Netherlands and helping me settle.

My amazing office mates: Abel Warries, Zoe Poirier Maruenda, Saskia Dunn, Dominique Loviscach, Tristan Chopinez, Louise Friis, Niels Schoubben, and Ziheng Cheng made the office a place I will miss. My PhD colleagues made

the daily work enjoyable and the long haul bearable: Djibrila Tetereou, Elizabeth Kerr, Bushra Alkhamees, Tina Bougag, Priscilla Lam, Bente De Graeve, Hamza Khwaja, Cid Swanenvleugel, Hadis Tamleh, Evgenia Mouresioti, Turan Hanci, Hester Groot, Melle Groen, Olga Nozdracheva, Sophia Nauta, Jiang Wu, Axel Palmér, Jingtian Shi, Nargess Asghari, Jesse Wichers Schreur, Cecilia Mignanti, Rasmus Thorsø, Federico Falletti, Pascale Eskes, Ariëlle Reitsema, Irina Morozova, Deborah Oyuu, Inocêncio Zandamela, Timothy Mac-Hadjah, Carmen Spiers, and others. My gratitude to Andrew Wigman for reading the manuscript with editorial care.

Old and new friends have been a vital support network: Djibrila Tetereou (also as a PhD colleague), Suleyman Nasir, Khalid M. Babikir, Jibreel Hessain, Al-Hasan Zaid, Walid Idris, Rihaam Abdelrahim, Khaled Altahir, Tagrid Ali Dinar, Mona Hashim, Hashim Badr, Tarig Homida, and Mustafa Basher. Rahel T. Dires deserves a special mention not only as the cover designer but for her steadfast emotional support during the lonely stretches of the PhD journey.

To all my colleagues at the University of Khartoum and particularly Abeer Bashir, Suzan Al-Amin, Helene Fatima Idris, and Maha Aldawi, I am grateful for your early encouragement and support. Rajend Mesthrie, my MA supervisor at Cape Town, has been a steady support for years.

Thanks also to the Labex LLACAN hosts and PhD colleagues in Paris for the productive research stay. I want to acknowledge the ERC Nilomorph project and Bert Remijsen for allowing me the time to finalize this PhD while starting my postdoctoral work.

I want to thank Bonny Sands, Mauro Tosco, Ed Elderkin, Giorgio Banti, Bert van Pinxteren, Morgan Nilsson, Marijn van Putten, Marwan Kilani, and my collaborator in the Sanye (Dahalo) documentation project, Kenneth Ngure. Thanks to the many language consultants in Tanzania and Kenya who shared words, stories, and local knowledge.

Family last, because they come first. My mother, Fatima, whose encouragement made schooling possible; my father, Sosal; my aunt Amina Abdalla; my siblings Hajir, Haliima, Babikir, Mohammed, Hawa, Aldheef, Maryam, Adam, Fathia, and Abulgasim; and the extended family — your steadiness and belief carried me through the hardest months. The COVID lockdowns and the (April 2023) war in Sudan were hard on us all; your support during those times was everything.

Ahmed Sosal
Leiden, 2026

1 Introduction

1.1 Background

Highland East Cushitic (HEC) is a subgroup within the East Cushitic branch of the Afroasiatic language family. Bender (1971, pp. 176, 222) introduced the term “Highland East Cushitic” to encompass Sidaama, Gedeo, Kambaata, T’ambaaro, Alaaba, Hadiyya, Libido, K’abeena, and Burji. This label replaced the earlier designation “Burji-Sidamo group” used by Moreno (1940), which itself derived from the traditional classification of the “Eastern Sidama group” (Sasse, 1982, p. 9). The set of phonological features that the HEC languages preserve provides insights into the sound system of their common ancestor, Proto-Highland East Cushitic (PHEC), including five vowels, contrastive vowel and consonant quantity, non-pulmonic consonants, and a phonemic glottal stop. This typological character makes the reconstruction of PHEC foundational for understanding phonological change across the broader Cushitic family.

In this dissertation, HEC is used as the cover label for the nine languages, while the comparative reconstruction sample is limited to five languages: Burji, Hadiyya, Kambaata, Sidaama, and Gedeo. This restriction is methodological and data-driven (in terms of comparability and documentation depth). The other named HEC varieties (Alaaba, K’abeena, Libido, Timbaro) are cited selectively as supplementary comparanda where relevant data are available.

The sociolinguistic setting of HEC is also relevant and needs a brief sketch before the literature review, data sources, research methods, and the organization of the study are addressed in the subsequent sections.

1.1.1 Number of Speakers

The available speaker statistics for the HEC languages span over three decades and reveal significant demographic changes. The table below presents estimates from various sources, with Hudson’s (1976) figures based on Bender (1971), Hudson’s (2007) data drawn from the 1994 Ethiopian census, and Eshetu Tadesse’s (2010) estimates using the 2007 Ethiopian census. Treis’ (2008) figures for Kambaata come from the 1998 population census.

Language	Speakers	Source
Burji	15,000	Hudson (1976)
	35,731	Hudson (2007)
	71,871	Eshetu Tadesse (2010)
Hadiyya	700,000	Hudson (1976)
	923,957	Hudson (2007)
	1,284,373	Eshetu Tadesse (2010)
Kambaata	250,000	Hudson (1976)
	487,654	Hudson (2007)
	606,241	Treis (2008)
	630,236	Eshetu Tadesse (2010)
Sidaama	850,000	Hudson (1976)
	1,876,329	Hudson (2007)
	2,966,474	Eshetu Tadesse (2010)
	Over 4,000,000	Yri & Pepper (2019)
Gedeo	250,000	Hudson (1976)
	637,082	Hudson (2007)
	986,977	Eshetu Tadesse (2010)

All five HEC languages show considerable growth in speaker numbers over this period. The figures for Sidaama jumped from Hudson's (1976) 850,000 to over 4,000,000 in Yri and Pepper (2019), a roughly fivefold increase. Part of this increase can be attributed to better census methods and to Ethiopia's population growth, which went from under 30 million in 1970 to over 100 million by 2019 (World Bank, 2026).

Hadiyya and Burji both exhibit demographic growth profiles consistent with this expansion. Hadiyya speaker estimates rose steadily from 700,000 to 923,957, reaching 1,284,373 by 2007. Burji remains the smallest HEC language community represented here, yet its figures likewise reflect significant growth, rising from 15,000 to 35,731, and reaching 71,871 by 2007.

The harmonious upward growth and combined total of several million speakers underscore the vitality of the HEC languages and emphasize the importance of this language group within Ethiopia's linguistic diversity.

1.1.2 Geographical Location

The four HEC languages, Hadiyya, Kambaata, Sidaama, and Gedeo, are spoken in Ethiopia. The main centre of Burji is Soyama, Ethiopia, and it has speaker communities in Moyale and Marsabit County, Northern Kenya. In Ethiopia, the HEC languages are found in what was formerly known as the Southern Nations, Nationalities, and Peoples Region (SNNPR), located in the southwest part of the country (see Figure 1). The speakers of each language reside in designated zones and weredas (districts) where the language serves as the primary means of communication (Eshetu Tadesse, 2010, p. 1).

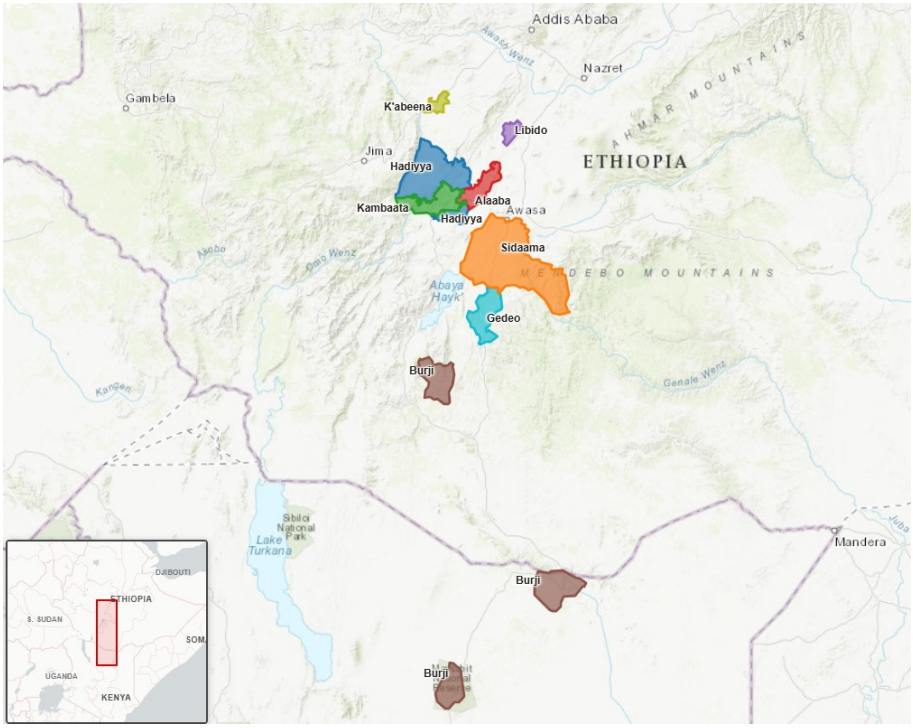


Figure 1: *Map of the Highland East Cushitic Languages*

Most of the HEC languages are spoken in contiguous territories in Ethiopia. The overall geographic proximity of the HEC communities has enabled sustained contact throughout their history, and this contact history shapes the distribution of shared innovations and contact-induced features of HEC phonology (see §1.5.4).

1.1.3 Classification

Over several decades of research, the classification of the HEC languages within the wider East Cushitic family has evolved considerably. The Language Survey of Ethiopia (1968–1970) established the early foundations for this work (Bender et al., 1976). As part of the survey, Fleming followed Moreno's (1940) earlier approach by treating HEC as a distinct subgroup within East Cushitic (Hudson, 1976, p. 246). Later studies by Wedekind (1980) and Hudson (1989, 2006) have refined our understanding of HEC phonology and lexicon.

Hudson (2006, pp. 295-296) identifies several phonological traits recurrent in the HEC languages: open syllables, a five-vowel system with contrastive length, consonant gemination, non-pulmonic consonants (e.g., **t'**, **k'**, **c'**, and in some cases **p'**), and a phonemic glottal stop. According to Hudson, these shared features, together with lexical and morphological evidence, support the coherence of HEC as a genetic grouping. In the present dissertation, however, these features are used as a descriptive profile rather than as standalone subgrouping diagnostics; the subgrouping argument is based on shared innovations and restricted distributions, developed cumulatively in the comparative analyses and synthesized in Chapter 5; a full Proto-East Cushitic innovation inventory lies beyond the scope of the present study.

The internal structure of Highland East Cushitic can be represented as follows:

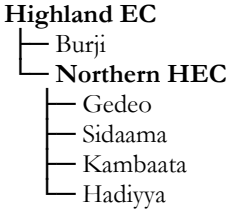


Figure 2: *Highland East Cushitic Family Tree*

The classification in Figure 2 shows a primary split between Burji and the Northern Highland East Cushitic (NHEC) languages. NHEC forms a cohesive subgroup characterized by shared innovations, as summarized in §5.3. Chapters 1–4 provide a detailed reconstruction of the phonological and morphophonological system of PHEC and clarify the historical changes that have yielded the above tree.

1.2 Literature and Data Sources

The East Cushitic lexical database (§1.3) contributes a comprehensive repository of lexical data from five Highland East Cushitic languages: Burji, Gedeo, Hadiyya, Kambaata, and Sidaama. These five are the primary analytical sample in this study. Other HEC languages are used only as auxiliary evidence in specific sections where such data strengthen or contextualize an argument. The data compiled from these languages reveals a complex interplay between the retained lexicon and subgroup-specific innovations. The fresh analytical tools and expanded data introduced by the database offer crucial insights into the phonological history of this East Cushitic branch.

The foundation for comparative Highland East Cushitic research remains Hudson's (1989) dictionary. It provides reconstructions for 215 (of 778) sets with attestations across four or more HEC languages. This pioneering work brought systematic comparison to HEC and remains the primary reference for comparative research. However, Hudson himself acknowledged critical methodological limitations. He "did not require regular correspondences for a reconstruction" and did not "[reconstruct] different proto-phonemes for different sets of correspondences", so his proto-forms function more as referencing devices than as fully substantiated reconstructions (Hudson, 1989, pp. 7–8). In this study, I build on Hudson's foundation while addressing these limitations through an expanded and updated comparative dataset, contextualized comparison within the broader Cushitic family, and implementation of explicit confidence evaluation for each reconstruction (see §1.4 and §1.5.1).

The HEC component of the database is characterized by its depth of lexical coverage, though representation varies across languages depending on the available sources. The documentation sources range from comprehensive dictionaries and grammars for Kambaata and Sidaama to more limited resources for Hadiyya and Gedeo. Such a state of uneven documentation is unavoidable and acknowledged in the evaluation system adopted in this study. This limitation does not fundamentally compromise the reconstruction, because regular correspondences can be established even with asymmetric data coverage when the patterns are recurrent across attested forms.

The following subsections detail the specific sources consulted for each language and explain how earlier work has been critically evaluated and integrated into this study.

1.2.1 Data Compilation and Sources

The Highland East Cushitic lexical data in the database (§1.3) comes from a selected combination of primary lexical resources and secondary grammatical sources. The starting point was Hudson's (1989) Highland East Cushitic dictionary, which remains the most comprehensive comparative dictionary for HEC and provides an essential baseline of lexical data. However, Hudson's work is now over three decades old, and the field has not stood still. His material has been systematically updated and expanded using more recent, language-specific sources to ensure the database reflects current scholarship and incorporates the considerable advances in documentation that have occurred since his dictionary was published.

For Burji, the database draws primarily on Roba Dame & Wedekind's (2008) unpublished dictionary, which provides updated lexical data and more refined transcriptions than earlier sources. The more extensive and recent Burji, Amharic, and English dictionary by Getahun Addo (2017) supplements Roba Dame & Wedekind's (2008) data. Sasse's (1982) etymological work remains valuable for its comparative insights, and Tesfaye Baye's (2015) grammar adds important morphophonological context that helps clarify the structure of lexical entries.

Kambaata benefits from Treis' wordlist, which includes recent fieldwork data, along with her comprehensive (2008) grammar that details the language's intricate morphophonological system. Alemu Banta's (2016) dictionary expands the lexical coverage beyond what was available in earlier sources.

For Hadiyya, Ritter's (2007) unpublished dictionary provides updated forms and glosses that correct and supplement earlier documentation. Tadesse Sibamo's (2015) grammatical description provides insights into the language's morphological structure, which informs the interpretation of the lexical data.

Sidaama is represented by Yri & Pepper's (2019) concise online dictionary, which is particularly useful for quick reference and current usage, along with Indrias Xa'miso et al.'s (2007) comprehensive Sidaama-Amharic-English trilingual dictionary, which remains the most extensive lexical resource for the language. The grammars by Anbessa Teferra (2000, 2014) and Kawachi (2007) provide essential grammatical context, particularly for understanding morphological patterns, which is crucial for interpreting their historical implications.

Gedeo lexical data comes primarily from Wedekind's (2008) unpublished dictionary, which is organized by semantic domain and includes careful attention to phonetic detail. Eyob Kelemework's (2015) and Gasparini's (1995) grammars supplement this with morphological analysis that clarifies structural patterns in the lexicon.

The grammatical descriptions are mainly used to supply the phonological processes at morpheme boundaries and morphological analyses needed later in this dissertation, especially in Chapter 3, where alternations at morpheme boundaries serve as central evidence.

The database also consults Lamberti & Sottile's (1997) study of Wolaytta, not because Wolaytta is part of HEC – it is Ometo, a separate branch of Omotic – but because their work includes valuable comparative data and illustrates language contact phenomena that help contextualize some of the patterns observed in the HEC languages.

This combination of data from Hudson's foundational comparative work with contemporary language-specific resources ensures that the database represents the current state of HEC documentation. Where Hudson's forms and glosses align with recent sources, this confirms the data's stability; where they differ, the more recent material typically reflects improved transcription practices, better understanding of phonological systems, or refinement of semantic analyses. The result is an HEC dataset that balances the systematic comparative perspective of Hudson's dictionary with the depth and accuracy of modern language-specific research. This dual foundation enhances both the reliability of the reconstructions presented here and the utility of this portion of the database for future research in Highland East Cushitic historical linguistics.

1.2.2 Previous Comparative Work

Hudson's (1989) dictionary represents the largest existing comparative work on Highland East Cushitic, but has significant limitations, as mentioned above. Those limitations, also discussed in reviews by Hayward (1997) and examined in Eshetu Tadesse's (2010) master's study, highlight the need for a more systematic approach grounded in explicit sound correspondences.

The foundations for HEC comparative work had been laid earlier. Wedekind's (1975, published 1980) phonological comparison of Sidaama, Gedeo, and Burji established some key correspondences within this subgroup, while Hudson's (1976) descriptive-comparative analysis extended the comparison

across all five HEC languages. Hudson's (2007) article on HEC morphology builds on this foundation, though he notes that some of his 1976 data had become outdated. Data from HEC also appears in broader East Cushitic comparative studies, including Black (1974), Sasse (1979), Ehret (1987, 1991), Arvanites (1991), and Bender (2020), as well as in work focused on related languages such as Bender's (1971) lexicostatistical study of Ethiopian languages and Lamberti & Sottile's (1997) Wolaytta comparative study.

What distinguishes the present study is its systematic approach to reconstruction. Rather than accepting Hudson's reconstructions as they are, I critically reassess his cognate sets, incorporate the expanded and updated lexical data described above, and apply explicit sound correspondences to evaluate each proposed proto-form. The confidence evaluation system introduced here (ranging from secure to tentative) addresses the uncertainties inherent in historical reconstruction. These evaluations help users assess which reconstructions rest on solid foundations and which remain more speculative, making the database a more transparent and practically useful tool for further research.

These literature sources feed into the broader East Cushitic Database, which situates HEC within its wider comparative context.

1.3 East Cushitic Database

The East Cushitic Lexical Database began as a personal research tool to organize the scattered lexical data across dozens of sources and bring some order to the historical complexities of the lexicon. What started as a spreadsheet has evolved into both a comprehensive Excel database and an interactive web application,[1] both of which are publicly available. While both versions contain the same core data, the web application is linked to specific data points (IDs) from this dissertation, allowing direct access to each lexical entry through its own URL, which leads to its fuller context within the database. This means that every reconstruction, every cognate set, and every linguistic claim made in this research can be traced back to its source with a simple click or by following the ID reference from the printed version.

[1] The Cushitic Lexical Database (CLD) web application is hosted on the GitHub Pages platform, which may change its hosting policies at any time. Hence, the long-term availability and sustainability of the web application version of the database cannot be guaranteed indefinitely.

Cushitic, as a branch of the larger Afroasiatic family, presents particular challenges for historical linguists. The internal diversity is striking, the documentation is uneven, and the linguistic history is layered with contact effects and innovations that can obscure older patterns. This database is my attempt to address these challenges by creating a structured framework for comparative analysis that permits us to track cognate sets, identify sound correspondences, and trace patterns of change across 23 East Cushitic languages. While coverage remains uneven, which is inevitable given the documentation gaps for some languages, this database represents the most comprehensive comparative East Cushitic lexical resource currently available. Future researchers will undoubtedly refine, correct, and expand upon this foundation. That is how the field advances.

1.3.1 Overview and Scope

At its core, this database is a comparative lexicon designed to support the reconstruction of Proto-Highland East Cushitic, though it encompasses data from across the East Cushitic family. The database's organization reveals how Highland East Cushitic patterns relate to the broader family, even if the primary focus remains on HEC itself.

Lexical coverage averages nearly 400 concepts per language, though this varies considerably depending on the state of documentation. This unevenness reflects the current state of Cushitic linguistics. Somali, for instance, is represented by around 750 entries, reflecting its relatively robust documentation. On the other end of the spectrum, languages like Yaaku, Ts'amakko, Gawwada, and Dahalo (Sanye) have fewer than 200 entries each. This reflects data availability, not linguistic importance. We work with what we have within the time available to us. In total, the database brings together over 8,500 compiled lexical entries drawn from more than 100 sources, including major dictionaries and grammars, as well as unpublished word lists and recent dissertations.

An important feature of the database is how it differentiates levels of reconstruction. All Highland East Cushitic and Northern Highland East Cushitic languages include proto-form reconstructions (marked as PHEC for Proto-Highland East Cushitic and PNHEC for Proto-Northern Highland East Cushitic). These reconstructions represent the core analytical work of this research. The other East Cushitic languages, including the Central (Oromoid), Northern, Omo-Tana, and Southern (Yaaku-Dullay) branches, are included for comparative purposes; however, no systematic reconstruction has been

attempted for these groups at this stage. Together with the Proto-West Rift (PWR) reconstructions, they provide context that helps identify broader patterns and serve as reference points for understanding the lexicon of HEC. This structure makes the database both a specialized tool for HEC research and a starting point for future comparative work across Cushitic.

For lexical reconstruction, a minimum attestation threshold is applied: a proto-form is reconstructed only when a cognate set is supported by at least three daughter-language cognates within the relevant scope. This is why some items have a PHEC reconstruction but no PNHEC reconstruction: a cognate set attested in Burji plus two NHEC languages meets the threshold for PHEC (HEC-wide), but does not meet it for PNHEC (NHEC-internal). Absence of a PNHEC reconstruction in such cases should be read as “insufficient internal evidence at this stage within the available data”, not as a claim that the item was absent from PNHEC.

This minimum cognate requirement is applied less strictly in morpheme boundary alternation processes and morphological reconstruction, where the relevant evidence may come from recurrent bound morphemes and paradigmatic alternations rather than from widely attested independent lexical items. Where reconstructions are based on narrower distributional evidence, they are treated as more tentative and are evaluated accordingly.

1.3.2 Organization and Presentation

Both the Excel database and the web application are organized around the same fundamental structure, though they present the data in different ways. The Excel version centers on a “Data” sheet where each of the 8,500+ entries occupies a single row, identified by a unique ID that represents a proposed cognate set. The 23 languages are arranged in columns according to a traditional subgrouping of the family:

Proto-Highland East Cushitic (PHEC):
- Burji
- Proto-Northern Highland East Cushitic (PNHEC): Hadiyya, Kambaata, Sidaama, Gedeo

Central East Cushitic (CEC):
- Oromo
- Sagan (S): Diraytata, Mositacha, Konso

Northern East Cushitic (NEC):
- Saho, Afar

Omo-Tana (OT):
- Bayso
- Eastern Omo-Tana (EOT): Rendille, Somali, Maay, Boni
- Western Omo-Tana (WOT): Arbore, Dhaasanac, Elmolo

Southern East Cushitic (SEC):
- Yaaku
- Dullay: Gawwada, Ts'amakko

These labels reflect revised East Cushitic subgrouping terminology; see §5.3.3 for detailed discussion and rationale. I have also included Dahalo, though its position within Cushitic remains debatable, and Proto-West Rift (PWR), which represents a broader Cushitic context. Their inclusion adds valuable comparative depth.

Each language's data follows a consistent structure. Take Burji as an example: Its columns include the lexical form itself (Burji), the gloss (BurGloss), grammatical information (BurMorph), contextual annotations (NoteBur), and bibliographic details (SourceBur). This same fourfold pattern of lexical form, gloss, morphology, notes, and source applies across all 23 languages in the database, ensuring systematic comparability while accommodating language-specific details.

Lexical Form: The primary word form as attested in the source
Gloss: The meaning(s) of the form, including multiple senses where relevant
Morphological Information: Details like case, gender, number, or verbal conjugations when available
Notes: Context annotations including dialectal variation, semantic nuances, and irregularities
Source Information: The publication and page number for each entry

The web application presents this same information through an interactive interface with features like filtering, sorting, and visualization tools. More importantly, it is accessible via hyperlinks on every entry ID in the PDF version of this dissertation. This means that every ID in this dissertation that references a particular cognate set or reconstruction has a direct link to that entry in the app, allowing readers to access all supporting data from each language, check the sources, and evaluate the evidence. This transparency is central to the research methodology behind this project.

1.3.3 Conventions and Visual Cues

The Excel version includes a “ReadME” sheet that explains the database’s conventions, transcription system, and methodological considerations. The web application incorporates this guidance directly into its interface via “About” and “Help” tabs in the header bar. They contain detailed keys for interpreting abbreviations (such as “CEC” for Central East Cushitic or “1SG.NOM” for first-person singular nominative). They also include a complete transcription guide that maps the database’s symbols to the International Phonetic Alphabet (IPA). This is particularly important because I have standardized transcriptions for some languages that use language-specific orthographic conventions in their published sources, such as Somali and Kambaata. The transcription guide ensures that researchers can verify phonological details and understand exactly what each symbol represents.

The Excel version uses color coding to make the data easier to navigate at a glance:

> **Red text:** language forms
> **Yellow text:** reconstructed proto-forms
> **Green text:** Oromo loanwords with a question mark where uncertain
> **Purple text:** Arabic/Swahili loanwords
> **Light blue text:** Amharic loanwords
> **Black-filled cells:** missing cognates, providing immediate visual cues for data gaps

The web application does not implement such visual distinctions.

1.3.4 Navigating the Database

Understanding the conventions in the Excel’s “ReadMe” sheet and the App’s “About” section is essential in order to interpret the data properly.

The database uses a consistent set of symbols throughout to indicate different types of information:

> **Semicolon “;”** separates entries from different sources, showing alternative attestations or updated data
> **Comma “,” or slash “/”** indicates variations within a single source (dialect differences, alternative transcriptions). Commas also separate multiple sources with identical data

Parentheses "( )" mark doubtful parallels where there are formal or semantic discrepancies[2]
Hyphen "-" explicitly indicates the absence of a cognate or parallel in that language
Empty cells are left blank for items I have not checked yet

A single cell often contains multiple forms or annotations. Here is an actual example from Burji:

Burji	BurGloss	BurMorph	NoteBur	SourceBur
líya; lia, liya	six	M	liyinka / liyinku 'the six'	Roba Dame & Wedekind (2008:57), Sasse (1982:135); Hudson (1989:135)

In this entry, *líya* is cited from Roba Dame & Wedekind (2008:57) and Sasse (1982:135). The comma between sources indicates they provide identical forms and glosses. The forms *lia, liya* come from Hudson (1989:135), representing alternative attestations from a different source; hence, they are preceded by a semicolon. The morphological information (M) comes from Roba Dame & Wedekind (2008:57). When notes draw from multiple sources, I cite the source directly within the note field.

1.3.5 Sources and Data Evaluation

The "Sources" sheet in the Excel version and the "Acknowledgments & Data Sources" section in the "About" page of the web application provide complete bibliographic information for every source consulted. This database draws on over 100 distinct sources, ranging from comprehensive dictionaries to unpublished data:

Dictionaries and Lexicons: Core vocabulary resources for individual languages, plus comparative lexicons from earlier scholars
Grammars: Grammatical descriptions that provide mainly morphological and syntactic context, often with a valuable wordlist or text appendices
Dissertations: Recent PhD and MA research with in-depth analyses and fresh data
Journal Articles and Monographs: Focused linguistic studies and comparative work

[2] Note that this has not been indicated consistently beyond HEC.

Unpublished Data: Lexical compilations including wordlists and dictionary drafts

For each source, I have compiled full bibliographic details: author, year, title, and publication information. This level of documentation serves several purposes. It allows for the tracing of any data point back to its origin, the assessment of the quality of transcription and glossing, and the situating of each lexical item within its broader research context. When multiple sources provide data for the same cognate set, this often reveals interesting variation, i.e., different transcriptions, different glosses, and sometimes different interpretations. These differences themselves can be linguistically significant.

This database builds on decades of documentation by numerous scholars. The dictionaries, word lists, grammars, and comparative studies they produced form the foundation of this work. Without their decades of documentation and analysis, none of this would be possible. The database is as much a tribute to their collective effort as it is a research tool.

An Excel version of the database is available on the LOT publications webshop.

1.3.6 Working Hypotheses and Data Interpretation

Another important point regarding the use and interpretation of the database is that the reconstructions presented for Proto-Highland East Cushitic (PHEC) and Proto-Northern Highland East Cushitic (PNHEC) are working hypotheses, not definitive conclusions. Historical linguistic reconstruction is inherently probabilistic and relies on making informed inferences based on available evidence, and besides that, different scholars may weigh evidence differently.

To reflect this reality, each reconstructed form is accompanied by a "Recon.Eval." confidence score ranging from 0–1 (evaluated as most secure) to 4 (evaluated as most tentative). These confidence scores serve as guideposts, helping users evaluate the evidence critically rather than taking every lexical reconstruction at face value (§1.5.1).

The confidence scoring system thus serves as an explicit acknowledgment that historical reconstruction operates on a spectrum of certainty, ranging from forms where regular correspondences, stable semantics, and broad attestation converge to produce high confidence, through intermediate cases where some evidence supports the reconstruction but complications remain,

to highly tentative proposals where the connection seems plausible enough to merit documentation but requires further investigation.

Given these inherent uncertainties, along with variation in source quality and transcription practices across such a large dataset, I strongly encourage a critical approach to using this database. Do not take everything at face value. I recommend the following:

> **Consult the original sources:** Use the detailed bibliographic information to track down the primary publications.
> **Cross-reference with other literature:** Look beyond the cited sources, as there may be more recent work or alternative analyses.
> **Assess the confidence scores:** Pay attention to which reconstructions are well-supported and which are more speculative.
> **Consider the context:** Look at related forms, check the broader cognate patterns, and think about the semantic plausibility.

Both the Excel database and the web application are designed to facilitate this kind of critical engagement. The application version makes it particularly easy to explore connections between entries, visualize language relationships, and trace the evidence trail for any particular claim. As mentioned earlier, every entry cited in this dissertation is cross-referenced and links directly to its full context in the app, making verification and deeper exploration straightforward.

Ultimately, this database serves as a research tool, designed to make comparative Cushitic linguistics more accessible and rigorous. It is not perfect – no database of this scale could be – but it is built on solid scholarship, organized systematically, and documented thoroughly. It represents the current state of our knowledge about East Cushitic lexical relationships, with an honest acknowledgment of where that knowledge is strong and where it remains uncertain. My hope is that it will serve not just as a resource for this dissertation, but as a foundation for future research that will refine, expand, and ultimately improve upon what has been compiled here.

1.4 Research Rationale and Objectives

The methodological framework outlined in §1.5 rests on a fundamental recognition that historical linguistic reconstruction operates on a spectrum of certainty. Not all reconstructions carry equal evidential weight, yet some traditional presentations often obscure this reality by treating all proto-forms

uniformly. This study addresses this issue through a probability-based evaluation framework applied to the expanded dataset described in §1.3.

Rather than treating all reconstructions as equally reliable, this approach assigns each proto-form an explicit confidence evaluation based on transparent criteria: regularity of sound correspondences, breadth of attestation, semantic uniformity, and absence of irregularities suggesting contact influence. This transparency provides a more honest representation of what we can confidently reconstruct about PHEC phonology versus what remains tentative or speculative, offering future researchers a clearer foundation for refinement or challenge.

The study pursues three interconnected objectives:

i. Reconstructing the phonological system of PHEC with explicit evaluation criteria. A probability scale classifies each reconstruction as secure, reasonable, arguable, or tentative (detailed in §1.5.1). This provides not just a picture of the PHEC sound system but an assessment of confidence in each reconstruction. This transparency distinguishes solid foundations from speculative proposals.

ii. Examining historical phonological changes through systematic correspondence analysis. The analysis traces how specific processes such as vowel doubling, consonant assimilation, and gemination have shaped modern HEC languages. Clear sound correspondences distinguish regular developments from analogical changes, contact effects, and sporadic innovations, providing a solid foundation for understanding subgrouping relationships (Chapter 5).

iii. Clarifying phonological processes at morpheme boundaries and developments at the phonology-morphology interface. Phonological changes interact with morphological structure in complex ways in HEC. Processes like epenthesis, nasal metathesis, and palatalization reveal where the two domains meet. Understanding why certain assimilations occur requires knowing both phonological environment and morphological context to determine whether we are dealing with causative suffixes, reflexive markers, or noun derivations. A comprehensive account of the phonological history of HEC must incorporate morphological analysis (Chapters 1 and 4).

These objectives are deeply interconnected. Phonological reconstructions inform sound change analysis, which helps interpret morpheme boundary alternations, feeding back into the evaluation of the reconstruction. The methodological framework below makes these connections explicit and systematic.

1.5 Methodological Framework

This study employs the historical-comparative method, reconstructing ancestral languages from attested daughter languages through systematic sound correspondences. The method rests on regular sound change: by identifying systematic correspondences between sounds in related languages, we can reconstruct proto-language phonology. The identification of morpheme boundaries and morphologically conditioned alternations is also necessary both for accurate comparison and for evaluating competing reconstructions. However, the application requires attention to data quality, explicit evaluation criteria, and transparent evidence presentation.

1.5.1 The Probability Scale for Reconstruction

A central methodological contribution is the four-tier probability scale for evaluating reconstructions. This builds on the confidence scores used for the reconstructed lexicon in the East Cushitic Database (§1.3.6) but adapts them into evaluation terms used throughout the dissertation: in Chapter 2 for phonological reconstruction, and where relevant in Chapters 3–4.

Not all reconstructions are equally reliable. Some rest on regular correspondences across multiple languages, clear semantics, and the absence of complications. Others exhibit irregular correspondences, limited attestation, or potential influence from contact. Rather than treating these as equivalent, the probability scale makes evidential differences explicit:

Secure Reconstruction (Equivalent to Score 0–1)

Supported by regular sound correspondences across all or most HEC languages, with no significant unexplained deviations. Stable semantics, neat formal fit, no indications of borrowing or analogical disturbance. In the database, a score of 0 is intended to serve as a more fine-grained filter for the few identical HEC cognates; otherwise, scores of 0–1 are equivalent on this scale.

Reasonable Reconstruction (Equivalent to Score 2)

Generally consistent with established correspondences, though one or two cognates may be missing or show minor irregularities. Irregularities are typically explained through independent sound changes, analogical leveling, or other processes. Solid working hypotheses supported by substantial evidence. Such reconstructions and those with scores of 0–1 form the backbone of the reconstructed lexicon (Chapter 2).

Arguable Reconstruction (Equivalent to Score 3)

Exhibits irregular correspondences or complications such as semantic drift, limited distribution, or competing explanations. Still within possibility and may represent genuine inherited items with unusual developments. Explicitly provisional, inviting further investigation.

Tentative/Provisional Reconstruction (Equivalent to Score 4)

Based on problematic correspondences requiring substantial additional evidence. May represent genuine cognates with highly irregular developments, ancient loanwords with obscured origins, or spurious comparisons. Included because even tentative reconstructions can identify potential patterns, highlight problem areas, and provide starting points for discussion.

This scale is systematically applied to every reconstruction, as indicated by **scores** in the database and **evaluation terms** in the subsequent chapters (2–3). They are interconnected because the extent of regularity in sound change, as discussed in the chapters, determines the regularity of lexical reconstructions, as shown in the database. This offers the possibility of evaluating evidence and distinguishing the solid core of PHEC phonology from more speculative elements.

The scale does not claim objective, quantifiable certainty, since historical linguistics does not work that way. It offers a consistent framework for making qualitative judgments about evidential strength explicit. Linguists might disagree on whether a reconstruction deserves a “2” or a “3”, but the scale provides a common vocabulary for discussion and forces explicit reasoning.

1.5.2 Terminology and Analytical Framework

To ensure consistency across analyses, the key technical terms are defined below:

Comparative Terminology

Comparanda or comparative set: Sets of forms selected for comparison across languages (each cognate set in §1.3). Sometimes used interchangeably with the “cognate set”.

Cognate set: Lexemes hypothesized to be inherited from a common ancestor.

Correspondence set/series: Systematic patterns of sound relationships across languages (e.g., Sidaama **t** : Kambaata **t** : Hadiyya **t** : Gedeo **t** : Burji **t** word-initially), whether inherited or due to contact.

Reflexes: Outcomes of proto-sounds in daughter languages.

Parallels: Forms that may represent loanwords, uncertain cognates, or false friends – in other words, potentially relevant but of uncertain status.

Borrowing/loanword: Terms used where lexical transfer from another language is supported by explicit comparative evidence (formal, distributional, or semantic).

Contact influence/contact-induced: Broad cover terms used where contact effects are likely, but the exact mechanism (e.g., direct borrowing, diffusion, analogical remodeling under contact) cannot be demonstrated with equal certainty. As a working heuristic, broader EC cognacy favors analyzing such cases as inherited material with contact-influenced reflexes (including possible double reflexes), whereas isolated irregular forms without wider EC support are treated as more likely borrowings.

Score: The numerical scores (0–4) (§1.5.1) assigned to assess the reliability of lexical reconstructions in the database.

Evaluation: The qualitative assessment terms (secure, reasonable, arguable, and tentative/provisional) used to evaluate reconstructions in

the study as descriptive equivalents of the numerical scores in the database (§1.5.1).

The study distinguishes between morphological markers and morphological/semantic functions. Hence, it uses different terms, as listed below.

Morphological Markers

Singulative: Markers deriving singular forms (***-ʃ**, often followed by terminal vowels) (§4.1.1.1)
Plurative: Markers deriving plural forms (e.g., ***-uw-a**) (§4.1.1.2)
Collective: Markers indicating group entities (***-ey-e**) (§4.1.1.2)
Case: Accusative and nominative forms (§4.1.5.2)
Passive: Marker ***-am**, consistent across all HEC languages (§4.2.1)
Reflexive: Markers ***-ʔ**, **-ak'k'**, **-it't'**, **-(V)d'** (§4.2.3)
Causative: Markers ***-s** and ***-i-s** with specific distribution (§4.2.2)
Deverbative: Deriving nominals from verbs (§4.1.2)
Denominative/Verbalizer: Deriving verbs from nominals (§4.2.5)

Morphological Functions

Singular/Plural: Number categories
Object/Subject: Grammatical relations
Middle: Subject of the verb is the principal interest of the action or state, e.g., acting on or for his or her benefit
Deverbal/Denominal: Derivational relationships between word classes

Phonological Terminology

Terminal vowel (TV): Default final vowel of nominal stems.

Stem-final vowel (V_i): Cover term to refer to terminal vowels and vowels from apocopated stems and fossilized suffixes.

Non-pulmonic consonants: Cover term for ejectives (**t'**, **k'**, **c'**) and implosive (**d'**), characterized by non-pulmonic airstream mechanisms. Mainly refers to consonants potentially inherited from a PHEC non-pulmonic consonant or borrowed consonants.

Glottalized consonants: Obstruents with secondary glottalization arising from (a) fusion of consonant + glottal stop (§3.3.1); (b) contact-induced

analogical developments (§2.1.2); or (c) assimilation to another non-pulmonic consonant. Distinguished from non-pulmonic consonants for historical analysis, though earlier literature (e.g., Sasse, 1979) often uses “glottalized” for both categories.

Palatalized consonants: Consonants that show a secondary palatalization resulting from (a) (morpho)phonological processes, (b) assimilation to a following or preceding palatal sound, or (c) sound symbolic palatalization process.

Palatal (or post-alveolar) consonants, on the other hand, refer to inherited palatal consonants, which are not products of the secondary processes associated with palatalized consonants (§3.2.1).

This framework provides a shared vocabulary for the subsequent chapters. When discussing “correspondence sets”, “reflexes”, or “middle”, these terms carry the specific comparative marking or function defined above.

1.5.3 Data Presentation and Referencing Conventions

This study utilizes consistent conventions for presenting lexical data, cross-referencing database entries, and formatting lexical examples. It is essential to understand these conventions to interpret the analyses in subsequent chapters.

Not all examples cited in the dissertation are drawn directly from the database. In particular, Chapters 3–4 frequently illustrate alternations with forms and paradigms and reconstruct suffixes cited from published grammars. Where an individual example or suffix is drawn from a grammar rather than a database entry, the source is cited in-text (author, year). Where a block of examples is presented as a set (e.g., suffixes, a paradigm, or a series of alternations), the sources are cited in a footnote for the block.

Examples and Referencing

Each cognate set from the database is assigned a unique ID, presented in square brackets next to the respective lexical examples (e.g., [240] *k’ort’umm-* ‘fish’) in the in-text examples (listed in paragraphs/bullet points); otherwise, they are placed above them in the block examples. There are multiple reasons for using these IDs:

Comprehensive data access: Every ID is a hyperlink directly to the corresponding database entry. Each set can be located in the database using the respective ID. There, readers can access all attested variants, full source bibliographies, morphological details, and comparative notes.

In-text referencing: IDs provide precise citations throughout the dissertation, accompanied by language abbreviations for clarity (e.g., PHEC [5] **san-* 'nose' → Burji *suna*; PNHEC **san-* 'nose' → Had *sane*, Kam *sanu-ta*, Sid, Ged *sano*). This system ensures that every reconstruction, every cognate set, and every claim can be traced back to its source evidence in the database.

Prioritization of recent data: The primary examples presented in-text are generally drawn from the most recent sources, reflecting current scholarship and improved transcription accuracy. However, when older forms, specifically those from Hudson (1989), offer greater accuracy or clarity based on comparative evidence, they are used instead, sometimes with explicit acknowledgment of variation in transcription.

Data Ordering and Presentation

The standard order of presentation for examples is PHEC, Burji; PNHEC, Hadiyya (Had), Kambaata (Kam), Sidaama (Sid), Gedeo (Ged). This sequence reflects the proposed subgrouping (detailed in §5) and facilitates the identification of innovations:

Burji patterns reveal the first split within HEC

Shared innovations among Hadiyya, Kambaata, Sidaama, and Gedeo establish Proto-Northern HEC (PNHEC)

Where appropriate for specific analytical purposes, this order may be adjusted to highlight particular points of discussion or to prioritize the language under focus.

When examples are identical across multiple languages, this is indicated by listing all the relevant language abbreviations before a single representation: "Had, Kam, Sid, Ged [461] *has-* 'to want, seek'" rather than repeating the same form.

Glossing Conventions

Simplified glosses are used for words with multiple meanings. Only the most relevant or primary gloss is included in block examples (e.g., PNHEC [70] **afoʔ-* ‘mouth’ rather than ‘mouth, opening, language’). The full range of attested glosses is available in the database entry.

Single glossing is used when the daughter language retains the reconstructed PHEC meaning or when the in-text examples have identical meanings. This means that individual glosses for all languages are omitted to avoid redundancy (e.g., PNHEC [14] **az-* ‘milk’ → Had, Sid, Ged *ado*, Kam *azu-ta*). Divergent glosses are provided explicitly (e.g., PHEC [316] **dam-* ‘branch, stick’ → Burji *dame* ‘branch’; PNHEC **dam-* ‘branch, stick’ → Kam *dama* ‘club, stick’, Sid *damatte* ‘branch’, Ged *dame* ‘branch’).

Bare **y** and **ʔ** in reconstructions denote the support verbs (§4.2.4) and verbalizers (§4.2.5), not suffixes. They appear without a hyphen when reconstructed in isolation or cited from a source, because they are independent verbs. A hyphen is used rarely where the form is attested as encliticized, that is, where the historical development of encliticization to the preceding word has produced a surface **-y** or **-ʔ** in the data. This convention prevents confusion with the glide/glottal suffixes discussed in Chapter 4.

Transcription and gloss variation are common. Variations in transcription (such as the presence or absence of a glottal stop before vowel-initial words) or glossing between sources reflect differences in analytical practice or fieldwork methodology. The database documents these variations; in-text examples use the most reliable transcription based on comparative evidence and recent scholarship. Significant variations are noted in discussions where relevant.

Transcription and Typographic Conventions

The following modifications have been applied to the transcription symbology of the data taken from different sources (modified = source transcription): **dʼ** = **ɗ**, **bʼ** = **ɓ**, **cʼ** = **tʃʼ**, **c** = **tʃ**, **y** = **j**, **j** = **dʒ**, **ʼm** = **ʔm**, **ʼn** = **ʔn**, **ʼl** = **ʔl**, **ʼr** = **ʔr**.

Several typographic conventions are used throughout the following chapters. Italic font is used for in-text examples discussed within paragraph prose (PHEC [237] **ʃirb* → PNHEC **sirb-* ‘to dance and sing’). Bold font highlights affixes (**-am**, **-is**), individual phonological segments (**tʼ**, **ʔ**, **w**), phonemes, and

occasionally the two support verb roots (***y*** 'to say, do, be' and ***ʔ*** 'to do') due to their short form. Square brackets [ ] enclose database IDs ([237], [14]) and phonetic transcriptions ([**w**]) when distinguished from phonemic representations. An asterisk marks reconstructed proto-forms (PHEC [72] **lam-* 'two').

These conventions ensure clarity, consistency, and transparency throughout the dissertation while facilitating direct engagement with the underlying database evidence.

1.5.4 Contact and Inheritance as a Continuum

A distinctive methodological feature of this study is the integration of contact and inheritance perspectives in phonological reconstruction. Rather than treating them as mutually exclusive explanations, i.e., inherited from PHEC versus borrowed through contact, this approach recognizes them as endpoints on a continuum that must be evaluated case-by-case based on distributional evidence. Additionally, the effect is not monotonous between the endpoints of the continuum.

This integrated approach is applied most systematically in the analysis of non-pulmonic alveolars and post-alveolar (§2.1.2), where irregular correspondences and distributional peculiarities call for careful evaluation of potential contact influence from neighboring languages and language families, particularly Omotic. For these segments, the analysis considers:

i. Regular correspondences to establish inherited phonemes through systematic patterns across HEC languages

ii. Distributional irregularities to flag potential contact influence, such as geographic clustering or lexical items with irregular reflexes

iii. Areal context from external comparison to Oromo, Omotic, and other neighboring families

Where correspondences are regular and there is evidence across the HEC languages, no systematic attempt has been made to tease apart potential contact from inheritance. The reason is that if such influence is present, it cannot be detected through available methods when patterns are fully regular within HEC. The absence of irregularity suggests either pure inheritance or contact so ancient and thorough that it has been regularized into the system.

In such cases, parsimony favors treating the pattern as inherited unless external evidence suggests otherwise.

The confidence scale system (§1.5.1) captures this complexity. Those reconstructions with regular correspondences but limited attestation or suspicious distributions are evaluated as reasonable, acknowledging that multiple historical processes may be at play. This approach avoids rigid binary classifications while maintaining analytical rigor.

1.6 Structure of the Dissertation

The dissertation is organized as an integrated study where chapters build upon and reinforce each other. Historical phonology, alternations at morpheme boundaries,[3] and morphology are interconnected aspects of linguistic structure, and the organization reflects these connections.

Chapter 2: Phonology

The chapter on phonology reconstructs PHEC consonant and vowel systems, syllable structure, and stress. Each proto-phoneme is analyzed for distribution, reflexes in daughter languages, and correspondence patterns justifying reconstruction. Consonants receive particular attention, especially the non-pulmonic series (ejectives, implosives), which present the most interesting historical questions. For these sounds, dedicated subsections address contact influence, particularly contact with the Omotic and Oromo languages.

Each section of the chapter presents correspondence patterns, exemplifying cognate sets, reconstructed forms, a discussion of complications, and further examples. Chapter 2 establishes the comprehensive PHEC sound system with an assessment of evidential strength.

Chapter 3: Alternations at Morpheme Boundaries

The chapter on phonological processes examines sound interactions at morpheme boundaries, including assimilation, palatalization, fusion, elision, metathesis, gemination, and epenthesis, and provides insight into the crucial interaction between phonology and morphology in HEC. Understanding these

[3] An example of alternations at morpheme boundaries is how Burji verb root like *jib-* 'run' changes when suffixes like **-s** (caus) are added, producing forms like *jiss-* through assimilation.

processes is essential because they explain surface alternations that might obscure historical relationships and provide additional evidence for lexical, phonological, and morphological reconstruction. The chapter also shows that tracing derivational history can turn apparently irregular forms into regular correspondences, thereby expanding the set of cognates available for reconstruction.

For example, suffix-triggered glottalization of preceding stops reveals both the phonological nature of the suffix and the phonological rules. Shared glottalization patterns across NHEC languages likely reflect inherited PNHEC morphophonological rules. Language-specific patterns, such as those in Burji, may represent innovations. Chapter 3 thus contributes to reconstruction by identifying old versus new boundary alternation patterns.

The chapter draws heavily on Chapter 2 for phonological foundations. For example, the discussions on glottalization require established glottal and other stop consonants and anticipate Chapter 4 by pinpointing morphological environments where these processes occur. This middle position makes Chapter 3 an essential bridge between the sound system and grammatical structure, although the choice of morphemes in Chapter 3 is mostly driven by their role in the sound change arguments developed in Chapter 2 and by the need to identify the morphological environments that condition those changes.

Chapter 4: Morphology

The morphology chapter reconstructs selected verbal and nominal morphemes, establishing the triggers of the boundary alternation processes described in Chapter 3. The focus is on the derivational suffixes and markers. While primarily concerned with verbal morphology (passive, reflexive, causative, and deverbative markers), a substantial portion treats nominal morphology (singulative, plurative, genitive, and case markers).

Morphology is included in this phonology-focused dissertation because in HEC languages, phonological history cannot be separated from morphological history. Most phonological changes are morphologically conditioned. Sound correspondences may be regular in roots but apparently irregular at word boundaries and in suffixes, or vice versa, blurring the cognate relationship. Understanding morphology is essential for accurately reconstructing phonology and explaining why sound correspondences sometimes break down regularity.

Chapter 4 also provides crucial evidence of subgrouping. Shared morphological innovations are among the strongest evidence for genetic relationships. The distribution of verbal extensions across the HEC languages plays a role in establishing family-internal structure.

Chapter 5: Conclusions

This chapter synthesizes the findings to address broader questions about the history and classification of HEC, organized around two themes: internal subgrouping (how the HEC languages relate to each other) and external contact (how interaction with neighboring languages, particularly Omotic and Oromo, has shaped HEC phonology and lexicon).

The subgrouping discussion draws on all three analytical chapters: shared phonological innovations (Chapter 2), shared boundary alternation processes (Chapter 3), and shared morphological markers (Chapter 4). Discussion about contact contextualizes irregular correspondences and exceptional patterns, showing how they often reflect borrowing, areal influence, or double-reflexes rather than irregular inheritance.

Chapter 5 also assesses implications for broader questions in Cushitic and Afroasiatic linguistics. What do HEC findings reveal about earlier East Cushitic stages? How do contact patterns illuminate the linguistic history of the Ethiopian highlands? What methodological lessons emerge for historical reconstruction in other families? These reflections connect detailed empirical work to larger issues in historical linguistics.

Appendix

The dissertation includes an appendix that provides an index of reconstructed lexical items cited throughout the study. Items are listed by their unique database identifier (ID) to facilitate cross-referencing, and each entry gives the PHEC and PNHEC reconstructions (where available) together with a gloss. This index is intended as a quick reference tool for readers to look up forms as they follow the arguments across chapters.

Navigating Cross-references

Frequent cross-references between chapters/sections (marked by §) are essential for following the arguments. Chapter 3 discusses glottalization, referencing Chapter 2 for relevant consonant reconstructions and Chapter 4 for triggering morphological environments. Chapter 4 analyses of causative

extensions reference Chapter 3 for boundary alternations and Chapter 2 for the established phonological pattern.

This web of cross-references reflects the reality that phonology, alternation processes, and morphology form an integrated system. Individual chapters provide focused information on specific topics, but the full picture emerges only when these chapters are understood in relation to one another. This dissertation is not an independent study, but rather a single, unified argument about HEC phonological history, presented through multiple complementary perspectives.

Major claims are supported by database cognate sets, identified by unique IDs and presented in example blocks. These IDs link directly to the web application (§1.3.2) for accessing full comparative data, checking sources, and evaluating evidence. This transparency is central to the methodology. Data and reasoning are provided for critical assessment.

1.7 Summary

This study attempts to advance our understanding of Highland East Cushitic phonological history through systematic comparison and transparent evaluation of the evidence. Historical reconstruction is not about discovering objective facts that we directly observe about proto-languages. Instead, it is about building the most coherent hypotheses we can from the available evidence. Although that does not exclude the possibility that the method can lead to objective conclusions. How strong the hypotheses are depends on data quality, methodological considerations, and an honest acknowledgment of uncertainty.

Hudson’s earlier work laid crucial groundwork, but three decades have passed. This study reassesses his proposals with new documentation, applies probability-based evaluation, proposes sound changes and their relative chronology, and examines how phonology, morpheme boundary alternations, and morphology interact. The result is a more robust reconstruction of PHEC. The probability scale makes it explicit that not all reconstructions carry equal weight. Acknowledging such variation strengthens the analysis because it allows readers to see which parts rest on solid ground and which do not.

Looking at phonology, boundary alternation processes, and morphology together while paying attention to language contact creates a model that recognizes both the regularity of sound change and the messy realities of

languages in contact. Sound correspondences are systematic, yet they are not mechanical laws. They interact with morphology, get disrupted by borrowing, and shift through analogy. Any complete historical account has to recognize all these factors.

The study also makes some broader methodological contributions. The probability scale provides a way to be explicit about the confidence in reconstruction. The inheritance and contact continuum demonstrates a reconstruction in a historically contiguous and dense language-contact context. The East Cushitic Database models how to organize comparative data so others can verify and extend the work. The integrated approach to phonology, morpheme boundary alternations, and morphology shows the value of treating these as interconnected rather than independent.

There is still plenty of work to be done. The database has tentative comparisons for non-HEC East Cushitic languages that need proper reconstruction. The boundary alternation processes in Chapter 3 deserve deeper investigation into their origin and patterns of spread. The contact phenomena in Chapter 5 raise larger questions about the linguistic history of the Ethiopian highlands and East Africa. The methodological approaches here could be useful for other Cushitic languages, other parts of Afroasiatic, or even other language families that deal with patchy data and complicated histories.

This dissertation ultimately contributes to an ongoing conversation among scholars working on Cushitic and Afroasiatic. It stands on decades of work by many researchers. My hope is that it gives future scholars something solid to build on, critique, and improve. That is how historical linguistics works, where each study answers some questions and raises others, each generation building on what came before.

Highland East Cushitic languages are spoken by millions of people in Ethiopia and Kenya. They are living, changing, vital parts of those communities. Understanding where they came from matters, not just academically but as part of the bigger effort to document and value linguistic diversity – this study is a contribution to that effort.

2 Phonology

This chapter reconstructs the phonological system of Proto-Highland East Cushitic (PHEC). To achieve this goal, the sound correspondences across five HEC languages (Burji, Hadiyya, Kambaata, Sidaama, and Gedeo) are analyzed by investigating the regularity of sound change. The reconstruction applies the probability-based evaluation framework established in §1.5. The confidence evaluation terms are used accordingly for each proto-phoneme based on the regularity of correspondences, the number of supporting examples, and hence the strength of comparative evidence.

Because reconstructed proto-phonemes (marked with *****) are used from the outset to formulate correspondence hypotheses, they should be treated here as working analytical labels. The full reconstructed phoneme inventory is presented in §2.5.1 (**Table *3***) and §2.5.3 (**Table *5***), and a consolidated class-based summary of the consonants is provided in Chapter 5 (**Table *10***).

This chapter demonstrates how the historical-comparative method reveals both inherited and contact-induced changes in the HEC phonological systems, as outlined in the methodological foundations in Chapter 1. The analysis follows an integrated approach that recognizes language contact as a pervasive factor shaping phonological change. Such an approach avoids the previous pitfalls of treating contact and inheritance as binary alternatives. Most of the sound correspondences support straightforward inheritance from PHEC. However, others are explained as a result of prolonged contact with neighboring languages, particularly Oromo. The reconstruction distinguishes between these scenarios by examining comparative evidence from the broader East Cushitic family where needed (see §2.1.2).

This chapter reconstructs the PHEC consonants, vowels, syllable structure, and accent. Section 2.1 establishes the consonant inventory of PHEC, examining pulmonic and non-pulmonic consonants across labial, alveolar, post-alveolar (palatal), velar, and glottal positions. Section 2.2 addresses the vowel system, including single, double, and terminal vowels. The two shorter sections that follow deal with the syllable structure (Section 2.3) and accent of PHEC (Section 2.4), respectively. The final Section 2.5 provides summaries and concluding remarks on the reconstructed phonology. The reconstruction evidence is presented systematically: the proposed proto-form, regular reflexes in the daughter languages, supporting examples from the database (with unique IDs in brackets), and a discussion and evaluation of correspondence regularity and reconstructed lexemes.

The terminological distinctions established in §1.5 are maintained throughout the reconstruction. Database entries are referenced with unique IDs in brackets (e.g., PHEC [72] **lam-* 'two'), enabling readers to examine full comparative sets in the database. Otherwise, the data source is cited. For further examples, only reconstructed lexemes are provided; readers can check the full context in the database by following the respective ID. Cross-references to other chapters (Ch. 3 for alternations at morpheme boundaries and Ch. 4 for morphology) appear where an apparently phonological question can be answered by morphology.

2.1 HEC Consonants

This section includes five subsections organized by manner or place of articulation:[4] pulmonic alveolars and post-alveolar (§2.1.1), non-pulmonic alveolars and post-alveolar (§2.1.2), labials (§2.1.3), glides (§2.1.4), and velars, glottals, and the pharyngeal (§2.1.5). The order is methodological rather than proportional. Subsections are arranged for analytical progression, and cross-references are used where related evidence and interpretation are developed in different sections. These subsections establish the consonant inventory of PHEC through the analysis of correspondence patterns. Each proto-consonant reconstruction is examined based on regularity, attestation, and comparative evidence, and evaluated using the probability-based framework (§1.5.1).

To establish the synchronic foundation for the diachronic reconstruction that follows, a brief overview of the consonant inventories attested across the modern HEC languages is provided in **Table *1***. Transcription symbols follow the conventions presented in §1.5.3.

The HEC languages collectively have 27 consonants, with individual languages ranging from 23 to 26 consonants.

[4] I have organized these subsections by manner and place to reflect patterns in the distribution of reflexes.

Table 1: *Highland East Cushitic Consonant Inventory*

	Labial	Alveolar	Post-alveolar	Velar	Glottal
Stop	**p, b**	**t, d**		**k, g**	**ʔ**
Non-pulmonic	**p', b'**	**t', d'**	**c'**	**k'**	
Fricative	**f**	**s, z**	**ʃ**		**h**
Affricate			**c, j**		
Lateral & Trill		**l, r**			
Nasal	**m**	**n**	**ɲ**		
Glide	**w**		**y**		

This study analyzes 24 of these consonants because Sidaama, Burji, and Kambaata **ɲ** and Burji **p**, **b'** are unattested in the database (§1.3). Several consonants require particular attention in reconstruction: the distribution of non-pulmonic consonants across languages, the variation in sibilant inventories (**ʃ** presence/absence), and the restricted attestation of certain phonemes (**ɲ**, **p**, **b'**). These synchronic patterns guide the identification of which consonant features are inherited versus innovative in each language.

As outlined in the chapter introduction, the reconstruction applies an integrated approach to contact and inheritance. This approach is particularly relevant for non-pulmonic consonants, where Oromo and Omotic influence is evident (§2.1.2). The reconstruction proceeds through comparative evidence from within HEC and, where relevant, broader East Cushitic. Proto-forms are documented with database IDs (e.g., PHEC [48] **tum-* 'to pound'), to enable cross-referencing to full comparative evidence.

2.1.1 Pulmonic Alveolars and Post-alveolar

This section analyzes the regular sound correspondences of Burji, Hadiyya, Kambaata, Sidaama, and Gedeo to reconstruct the pulmonic alveolar and post-alveolar consonants in PHEC. While the stops ***t** and ***d** are securely reconstructed with consistent reflexes, the fricatives present more complex patterns. The voiced fricative ***z** has merged with ***d** in most languages, and the voiceless alveolar fricatives ***s** and ***ʃ** exhibit sporadic retention alongside merger patterns. Confidence evaluations (secure, reasonable, arguable, or tentative) for these reconstructions follow the framework outlined in §1.5.1. This section examines stops (§2.1.1.1), fricatives (§2.1.1.2), laterals and trills (§2.1.1.3), and nasals (§2.1.1.4), establishing secure reconstructions for ***t**, ***d**, ***l**, ***r**, ***s**, and ***n**, while the fricatives ***z** and ***ʃ** present more complex patterns requiring careful evaluation. Additionally, the phonological processes

outlined in the phonological processes chapter (Ch. 3), such as assimilation and palatalization, are essential for interpreting the reflexes of the common suffixes consisting of alveolars and post-alveolar.

2.1.1.1 Stops

The alveolar stops ***t** and ***d** reconstruct to PHEC, with regular correspondences across all five daughter languages.[5]

PHEC *t

The reconstruction of the voiceless alveolar/dental stop ***t** in PHEC is supported by regular reflexes across the HEC languages. Reflexes of ***t** appear in both initial and non-initial positions, as illustrated by the following cognate sets:

	[48]		[521]		[605]	
PHEC	***tum-**	'to pound'	***it-**	'to eat'	***galat-**	'thanks'
Burji	**tum-**	'to pound'	**it-**	'to eat'	**galataa**	'thanks'
PNHEC	***tum-**	'to pound'	***it-**	'to eat'	***galat-**	'thanks'
Ged	**tum-**	'to pound'	**it-**	'to eat'	**galata**	'thanks'
Sid	**tum-**	'hit, strike'	**it-**	'to eat'	**galata**	'thanks'
Had	**tum-**	'to pound'	**it-**	'to eat'	**(galata)**[6]	'thanks'
Kam	**tum-**	'to pound'	**it-**	'to eat'	**galata**	'thanks'

These reflexes illustrate the stability of ***t** across HEC languages. However, alterations involving **t** at morpheme boundaries reveal processes such as assimilation, gemination, and, particularly in Burji, palatalization – see §3.1 for alternation processes affecting alveolars.

Further examples supporting the reconstruction of PHEC ***t** include: PHEC [40] **tuf-* 'to spit', [130] **tuul-* 'to pile up', [71] **k'utaħ-* 'thorn', [87] **k'ot-* 'to dig', [180] **utub-* 'pole, center-pole of house', [142] **tam-n-* 'ten', and PNHEC [254] **tun~tumm-* (← **tum~tum-)* 'fist', PHEC [462] **meet-* 'scraping board for use in preparing ensete', and PNHEC [460] **beet-* 'son, child'.

Overall, the reflexes of PHEC ***t** across HEC languages demonstrate its stability as a phoneme in both initial and non-initial positions, and it can be

[5] According to Sasse (1982, p. 16), Burji **t** and **d** are dental [t̪] and [d̪] rather than alveolars. This might be an areal pronunciation reported for Rendille, Dahalo, and coastal Bantu languages as well (see Nurse, 1985).

[6] The parentheses indicate a problematic parallel (§1.3.3 and §2.1.1.3), and the Hadiyya reflex **l** is irregular in this example.

reconstructed as a secure proto-phoneme. However, Burji's distinct palatalization (§3.1.2.1) introduces an innovative layer to this phonological picture.

PHEC *d

Parallel to ***t**, the voiced alveolar stop ***d** is also securely reconstructed for PHEC. The reflexes are consistent in initial and non-initial positions.

	[672]		[325]		[378]	
PHEC	***dul-**	'to skin/ slaughter'	***duud-**	'mute (person)'	***dan-**	'be able'
Burji	**dul-**	'pierce'	**duuda**	'dumb'	**dan-**	'be able'
PNHEC	***dul-**	'to skin/ slaughter'	***duud-**	'mute (person)'	***dan-t-**	'be able'
Had	**dur-**	'slaughter, kill'	**duuda**	'mute'	**-**	-
Kam	**dul-**	'slaughter'	**duuda**	'mute'	**dand-**	'be able'
Sid	**dul-**	'skin, flay'	**duuda**	'mute'	**danda?**	'be able'
Ged	**dul-**	'skin'	**duuda**	'blunt, dull'	**dand-e?-**	'be able'

The above reflexes reveal a retention of ***d**, particularly in the initial position. In the case of [378] **dan-* 'be able', the absence of the final **d** in Burji compared to the PNHEC form **dan-t-* is due to a fossilized denominative suffix **-t**, which underwent voicing assimilation as discussed in §3.1.1.1. The verbalizer **-ʔ** in Sidaama and Gedeo (§4.2.5.1), and potentially the reflexive in Burji *dan-ʔ-* 'able for oneself' (§4.2.3.1), indicates later independent suffixation.[7]

Further examples supporting the reconstruction of PHEC ***d** include PHEC [316] **dam-* 'branch, stick', [114] **debeʔ-* 'to belch', and [206] **awd-* 'threshing floor', PNHEC [389] **duk-* 'to carry, to creep', [629] **bad-* 'to mistreat', [606] **gubeed-* 'thigh', [399] **dawam-* 'red', and [436] **ad-* 'aunt'.

The reflexes of ***d** in HEC underscore its stability across HEC. For the relevant alternations involving some non-initial reflexes of ***d**, see §3.3.1.2.

2.1.1.2 Fricatives

The reconstruction of fricatives in PHEC presents a more complex picture than the stops, particularly with regard to the voiceless alveolar fricatives. The

[7] see Gawaada *tanaʔ-* and Oromo *dandaya* 'be able', probably with denominatives as well (§3.1.1.1).

fricative system poses methodological challenges: distinguishing merger from retention, evaluating contact effects, and assessing the reliability of isolated cognates. The analysis below prioritizes systematic correspondences over sporadic attestations.

PHEC *z

The reconstruction of ***z** rests on regular correspondences in Kambaata and Alaaba, contrasting with the ***z** → **d** merger in Burji, Hadiyya, Sidaama, and Gedeo. The following comparanda illustrate both the retention pattern and the contrast with **d** reflexes:

	[118]		[14]		[139]	
PHEC	***zaak-**	'to swim'	***az-**	'milk'	***wozan-**	'heart'
Burji	**daah-**	'swim'	**ada**	'milk'	**wodana**	'heart'
PNHEC	***zaaz-, *zaak**	'to swim, flow'	***az-**	'milk'	***wozan-**	'heart'
Had	**(daad-)**	'spill, flow'	**ado**	'milk'	**wodano**	'heart'
Kam	**(zaaz-)**	'spill, flow'	**azu-ta**	'milk'	**wozana**	'heart'
Sid	**daak-i**	'swim'	**ʔado**	'milk'	**wodana**	'heart'
Ged	**daak-**	'swim'	**ado**	'milk'	**(onna)**	'heart'

In [118], the crucial evidence for ***z** is Alaaba ***zak-*** 'swim', which simultaneously preserves the ***z** reflex (parallel to Kambaata) and the 'swim' meaning (parallel to Burji/Sidaama/Gedeo). The Kambaata and Hadiyya stem-final consonant in *zaaz-* / *daad-* 'spill, flow' may reflect assimilation following the lenition of **k** to **h** (§3.2.2). Semantically, a different Sidaama lexeme, *daad-* 'spill, flow' (cf. Gedeo *dan-am-* 'spill, flow'), aligns better with the Hadiyya and Kambaata meaning, unless borrowed from Hadiyya/Kambaata.[8] Without Alaaba, the case for [118] would be weak; with it, the ***z** reflex and the 'swim' gloss are linked in a single witness, and the strongest support for PHEC ***z** then comes from [14] *az-* 'milk' and [139] **wozan-* 'heart', where the correspondences are straightforward. In [139], Hadiyya shows variation between *wodano* and *wadano* 'heart', while Gedeo *onna*, without the initial **w**, is likely borrowed from Oromo *onn-ee* 'heart'.

Despite the limited retention of ***z** in only Kambaata and Alaaba, both closely related languages within NHEC, the broader EC comparanda (e.g., Arbore, Saho **z**, and PHEC ***z** correspondences) support the reasonable

[8] Oromo *daaka* 'swim' provides a further indication of the possibility that the Burji, Gedeo, and Sidaama parallels are borrowed.

reconstruction of ***z** to PHEC, with a subsequent merger in most daughter languages.

The retention of ***z** in Kambaata and the related language provides crucial evidence for reconstructing it for PHEC. However, the widespread merger with **d** in Burji, Hadiyya, Sidaama, and Gedeo requires an explanation. Two competing analyses have been proposed to account for the distribution. Hudson (1989, p. 8) proposes reconstructing ***dz** for PHEC, arguing that it simplifies the derivation: ***dz** → **z** in Kambaata/Alaaba and **dz** → **d** elsewhere. Sasse's (1976, 1979) simpler ***z** reconstruction is preferred here. Hudson's derivation is also parsimonious, but ***z** accounts for the Kambaata/Alaaba reflex directly, i.e., no deaffrication step is required and is consistent with the absence of affricates elsewhere in the reconstructed PHEC inventory. Hudson's ***dz**, by contrast, introduces a proto-segment with no other consonant classification support.

Further examples of PHEC ***z** reflexes include: PHEC [438] **zuug-* 'to scrape, tan (hide)', [511] **zal~zal-* 'trade, business', [581] **zar-* 'to tear', [330] **sooz-* 'dawn', [383] **orz-* 'hair of animals', and [444] **bagaz-* 'spear'.

HEC languages other than Kambaata and Alaaba do have instances of **z**, but only in loanwords. For example, Anbessa Teferra (2000, p. 13) notes that **z** in Sidaama occurs only in Amharic loans, and Hudson (1976, p. 248) similarly observes its rare attestation in Hadiyya and Gedeo except in borrowed words. Bender (2020, p. 116) also notes that **z** is found rarely in Burji and is restricted to loans.

The comparative evidence supports a reconstruction of PHEC ***z** with limited retention and widespread merger with **d**. This aligns with a broader pattern in East Cushitic languages, where only a handful of languages retain a reflex of **z** against the majority that show a corresponding **d**.

PHEC *s

The voiceless alveolar fricative ***s** reconstructs securely to PHEC based on consistent reflexes across all five languages in both initial and non-initial positions.

	[5]		[117]		[7]	
PHEC	***san-**	‘nose’	***saʔ-**	‘cow’	***is-u**	‘he’ (3sg.m.nom)
Burji	**suna**	‘nose’	**saa**	‘cow’	**ʔísị**	‘he, it’
PNHEC	***san-**	‘nose’	***saʔ-**	‘cow’	***is-u**	‘he’ (3sg.m.nom)
Had	**sane**	‘nose’	**saáyyạ**	‘cow’	**isse**	‘he’ (honorific)
Kam	**sanu-ta**	‘nose’	**saʔa**	‘cow’	**ís/ísu**	‘he’
Sid	**sano**	‘nose’	**saa**	‘cow’	**ʔísi**	‘he’
Ged	**sano**	‘nose’	**saayya**	‘cow’	**isi**	‘he’

Further examples include: PHEC [723] **sek-* → **seh-* (→ **seh-s* (caus)) ‘adorn, decorate’, [20] **sazeħ-* ‘three’, [39] **suk-ʔ-* ‘to spin’, [637] **sim-* ‘ensete tree 1st stage’, [330] **sooz-* ‘dawn’, [108] **sunk’-* ‘to kiss’, [638] **sor-* ‘red soil’, [350] **sang-* ‘ox, bull’ (see Amharic *sänga*, Oromo *sangaa* ‘ox for beef’), [461] **has-* ‘to want, seek’, [633] **has y-s-* ‘be necessary’, PNHEC [601] **hos-* ‘to pass/spend the day’, PHEC [603] **boos-* ‘fireplace’, and PNHEC [33] **osal-ʔ-* ‘to laugh’.

The reconstruction of ***s** for PHEC is secure, supported by consistent correspondences across all five HEC languages.

Sporadic cases of ***s** → **f** show Oromo influence on Burji because it is a typical Oromo sound change. Such an influence is attested in two HEC numerals, though these words with **f** are not attested in Oromo. The shift is not phonologically conditioned and cannot be predicted from the environment. It is item-specific, contact-induced replacement reflecting the broader Oromo influence on Burji rather than a Burji-internal sound change. PHEC **s** otherwise remains **s** in Burji.

> PHEC [20] **sazeħ-* ‘three’ → Burji *fadiya*, PNHEC **saseħ-* ‘three’ → Had *saso*, Kam *sasu*, Ged and Sid *sase*.

> PHEC [467] **ħons-* ‘nine’ → Burji *wonf-a*; PNHEC **ħons-* ‘nine’ → Had *honso*, Kam *honsu*, Sid *honse*.

There is one probable instance of hypercorrection reverse change ***f** → **s** in Burji:

> PHEC [120] **fil-ʔ-* ‘to comb’ → Burji *fil-, sil-*; PNHEC **fil-ʔ-* ‘to comb’ → Had *fit’-*, Sid *fit’t’-*, Kam *fiil-*, Ged *fil-*.

Sasse (1982, p. 166) argues that Burji *sil-ʔ-* ‘to comb’ is unrelated to Oromo *fil-* ‘to comb’, which has original **f** from PEC **fil-* ‘separate’; his account implies

Burji preserves only the sibilant form. However, Getahun Addo (2017, p. 307) documents Burji *fil-* ‘to comb’ with the expected regular **f** reflex of PHEC and PEC, coexisting with *sil-ʔ-* in the same language. The presence of both forms in Burji is what undermines Sasse’s separation and supports the hypercorrection reading taken here: *fil-* is the inherited form, and *sil-ʔ-* is a Burji-internal hypercorrection back to ***s*** under sustained Oromo contact.

PHEC *ʃ

The evidence for reconstructing PHEC ***ʃ*** is more tenuous than that for other fricatives, paralleling the situation in the rest of East Cushitic (see Sasse, 1979, p. 31). Sporadic initial and non-initial correspondences are provisionally taken as evidence for traces of retained and re-established PHEC *ʃ via innovated and borrowed lexicon with no clear East Cushitic cognates (see §2.1.5.1 (PHEC *k’)).

The analysis below progresses through these dimensions: first establishing the basic correspondence sets that justify the reconstructions, then addressing the distributional and irregular patterns and their explanations, followed by systematic treatment of contact phenomena, areal effects, and finally geminate attestations.

	[8]		[497]		[249]	
PHEC	***ʃool-**[9]	‘four’	***ʃakol-, *ʃol-**	‘be easy, light’	**-**	-
Burji	**foola**	‘four’	**(ʃakk-ad’-)**	‘be easy’	**-**	-
PNHEC	***ʃool-**	‘four’	***ʃoll ʔ**	‘be easy’	***ʃik’ y / ʔ**	‘to approach’
Had	**sooro**	‘four’	**ʃoll-eʔ-**	‘be easy’	**ʃik’aaʔ-**	‘approach’
Kam	**ʃoolu**	‘four’	**ʃalal-**	‘be easy’	**ʃik’ y**	‘budge’
Sid	**ʃoole**	‘four’	**ʃot-**	‘be light’	**ʃik’k’iʃi**	‘approach’
Ged	**ʃoole**	‘four’	**ʃolled’-**	‘be light’	**ʃik’k’a**	‘approach’

In PHEC [8] **ʃool-* ‘four’, Oromo influence in the form of ***s*** → **f** is evident in Burji *foola*, despite Oromo having [1035] *afur* ‘four’. Burji **f** is not a direct reflex of PHEC ***ʃ***; it traces via the chain **ʃ** → **s** → **f**, where the second step is the Oromo-influenced **s** → **f** shift already established under PHEC ***s*** above. The Burji **s** ~ **ʃ** alternation documented across these sets is what allows the intermediate stage to remain visible; the **f** output is the contact-induced terminus of the same chain. Hadiyya’s **s** reflex is unique in HEC, so it might

[9] PHEC [8] **ʃool-* ‘four’ via an intermediate **ʃoħl-* if related to PD **salaħ* ‘four’ both from earlier **salħ*? Vowel quality is problematic.

be an irregular change.[10] Examples of the verbalizers like *ʔ* and the other reflexes of the denominative in [497] are discussed in §4.2.5.[11] Further examples of initial **ʃ** include: PHEC [95] *ʃomb-* 'lung' and PHEC [237] **ʃirb* 'to dance and sing'.

The reconstruction of ***ʃ** for PHEC is arguable, given the complex merger patterns, its sporadic retention, and the areal influences documented below.

All arguable retentions of PHEC ***ʃ** in the present dataset are word-initial. In intervocalic (non-initial) position, there is no direct reflex contrast between inherited ***s** and ***ʃ** in HEC, which points to an extensiveness of the merger in this environment.

> PHEC [252] **hasaaw-* (← **haʃaaw-*) 'speak, talk' → Burji *hasaaw-* 'to talk'; PNHEC **hasaaw-* 'speak, talk' → Had *hasaaw-* 'chat', Kam *haasaaww-* 'discuss', Sid *hasaaw-* 'speak, chat, talk', Ged *haaso?* 'speak, chat'.

Set [252] is therefore treated as an indirect case: the non-initial **s** can be interpreted as the outcome of earlier ***ʃ** only by comparison with set [229] PHEC **haʃ~haʃ-* 'to whisper'. The reconstruction **haʃaaw-* is inferential, not based on direct retention of a non-initial ***ʃ** contrast, and any semantic link between the two sets is treated as tentative. An additional comparative indication comes from Oromo: since PEC ***s** regularly yields Oromo **f** (Sasse, 1979, p. 33), the **s** in Oromo cognate *haasa(a)y-*, *aasaay-* 'speak, chat' and *(h)asaasa* 'to whisper' cannot derive from PEC ***s** and is better traced to PEC ***ʃ**, which independently supports the reconstruction of (Pre-PHEC) **haʃaaw-*. Burji **s** (instead of expected **ʃ**) thus reflects leveling/contact pressure from the same Oromo form rather than direct inheritance of ***s**.

For some instances of non-initial **ʃ** adjacent to consonants, it is difficult to determine if they are part of the root or suffixed allomorphs of the causative (see §3.1.2.3) or causative denominative (§4.2.5.7) because they are almost exclusively found on verbs. An example is the likely metathesized suffix **-s**/**-ʃ** in Burji [45] *fiisk'-* and Hadiyya *fiʃkaʔa* (intr) 'to whistle'. The reflexes of **ʃ**

[10] Similar **s ~ ʃ** variation in NHEC appears in the Amharic loanword [394] *ʃənkurt* 'onion' → Burji *ʃunkurtee*, Ged *ʃunkurte*, Sid *sunkurta*, Had *ʃunkurutta* / *ʃinkurutta* / *sunkurutta*, Kam *sunkutta-ta* / *ʃunkurta-ta*. Hadiyya and Kambaata show free variation between **ʃ ~ s**, of the Amharic **ʃ**, while Sidaama shows only **s**.

[11] On the notation of bare y and ʔ (e.g., **ʃik' y* / *ʔ* 'to approach') as support verbs and not as suffixes, see §1.5.

adjacent to consonants can be approximately generalized as NHEC **ʃ**, and Burji **s** or **ʃ**.

> PNHEC [693] **darʃ-* ‘to swell’ → Had *daʃʃ-* ‘swell’, Sid, Ged *darʃ-*.
>
> PHEC [229] **haʃ~haʃ-* ‘to whisper’ → Burji *hasaas-* ‘whisper’; PNHEC **haʃ~haʃ-* ‘to whisper’ → Had *haʃaʃa wac-*, Kam *haʃaaʃʃ-*, Sid *haʃaaʃ-*, Ged *haʃaʃ-*.
>
> PHEC [109] **ʃ-* ‘to kill’ → Burji *siy-*; PNHEC **-ʃ-* ‘to kill’ → Had, Kam *ʃ-*, Sid *ʃi*, Ged *ʃi-ya*.

Again, the Burji word is closer to Oromo *(h)asaasa* ‘whisper’ in [229], which might explain the **s** reflex. The full reduplication is reconstructed provisionally to account for the double V_2 reflexes attested in Burji, Kambaata, and Sidaama (§2.2.2.2), which would otherwise be unmotivated under a partially reduplicated proto-form, once the weak non-initial **h** and the absence of **h** in clusters are taken into account. This does not deny the onomatopoeic character of the root but treats it as having stabilized into a regular reduplicated stem at the proto-level, which is why it remains a usable witness here.

The EC parallels of PHEC [109] **ʃ-* ‘to kill’ show **ʃ** in Konso *ikaʃ-*, *iʃʃ-* ‘kill’, **f** in Saho *igdife* ‘to kill’, and **s** in Oromo *ajees-*, *ijees-* ‘kill’, Rendille *agiis* ‘to kill’, and Elmolo *ékis* ‘to kill, extinguish’. Sasse (1982, p. 167) reconstructed this root as a prefix verb in PEC **-gʃi* ‘to kill’. If PEC had **-gʃi*, that means PHEC ***ʃ** became Burji **s** via ***g** devoicing to ***kʃ** and subsequent lenition to ***hʃ**, followed by the loss of the fricative **h** (§2.1.5.2). Burji *siy-* and Gedeo *ʃi-ya* may be compounds including the support verb **y* ‘to say, do, be’. Kambaata also has *ʃiyye* (imp.pl) ‘kill!’.

Proto-East Cushitic ***ʃ** and ***s** regularly became Oromo **s** and **f** and PNHEC ***ʃ** and ***s**, respectively. The NHEC reflexes are sporadic at times, so Oromo reflexes are used to determine some proto-phonemes, such as the following:

> Oro *sama* ‘to rot, mould’, PHEC [523] **ʃam-* ‘to rot, to be wet’ → Burji *sam-* (from Oro), *ʃam-*; PNHEC **ʃam-* ‘to rot, be wet’ → Had, Kam, Ged, and Sid *ʃam-*.
>
> Oro *saalf-ad’d’a* ‘be shy, be ashamed’, PHEC [450] **ʃaal-* ‘be shy, be ashamed’ → Burji *ʃa(a)lʔ-* ‘be shy’, *saalfad’-* ‘be embarrassed’;

PNHEC **saal-* ‘be shy, be ashamed’ → Kam *saal-*, Sid *saal-*, *saalfat-*, Ged *saalfat-*.

Oro *finc’aan* ‘urinate’, PHEC [170] **sum ʔ* ‘urinate’ → Burji *sinʔ-aaw-*, PNHEC **ʃum ʔ* ‘urinate’ → Had *ʃum-eʔ-*, Kam *ʃum-aʔ-*, Sid *ʃumi*, *finc’i*, Ged *ʃiʔn-*.

Oro *reefa* ‘corpse’, PHEC [329] **reħ-ʃ-* ‘corpse’ (see [184] **reħ-* ‘die’) → Burji *reesi, reefa* ‘corpse’; PNHEC **reħ-ʃ-* ‘corpse’ → Had *leeʃa*, *leeʃʃa* ‘corpse’, Kam *reʃa*, Sid r*eeʃʃa*, Ged *reenʃa*.

PHEC [450] **ʃaal-* ‘be shy, be ashamed’ is mainly supported by Burji and Oromo reflexes, and so is reconstructed assuming that they are strictly regular. Burji, Sidaama, and Gedeo *saalf-* ‘be shy, be ashamed’ is from or influenced by Oromo *saalf-ad’d’a* ‘be shy, be ashamed’ (§2.1.1.2).[12] However, PHEC [170] **sum ʔ* ‘urinate’ is challenging because Burji shows irregular reflexes, with **ʃ** found on the noun *ʃinʔaa* ‘urine’ (see [145]) and **s** on the verb *sinʔ-aaw-* ‘urinate’ (see Amharic *ʃänna* ‘urinate’). Getahun Addo (2017, p. 631) documented Burji *sinʔa* ‘urine’ with a regular **s**. The noun [329] **reħ-ʃ-* ‘corpse’ is derived via the derivation suffix **-ʃ** (§ 4.1.1) and challenges the full reliance on Oromo since it is an example of skewed **f** reflex via **ʃ** and **s**, not a direct ***s** (see Ehret, 1987, pp. 139–141).

The following discussion is not an account of any single set already presented; it covers a transversal **s ~ ʃ** pattern that crosscuts the inheritance and contact analyses of §2.1.1.2 and connects to the broader obstruent-harmony hypothesis developed in fn. 36. Hayward’s (1988, 1997) sibilant-harmony framework, introduced in Hayward (1988, pp. 274ff) and described in Hayward (1997, pp. 115–118) as an Omotic areal feature adopted by early HEC speakers, provides an arguable account of certain irregularities. In many Omotic languages, sibilant obstruents organize into “palatal” and “non-palatal” sets, maintaining strict segregation within roots. The change of ***t’*** → **c’** in Burji *c’eeji* (← PHEC [18] **t’iig-* ‘blood’) likely reflects harmony triggered by the palatalized stem-final ***j*** (← ***g***) (§3.2.1, fn. 36). The analogous obstruent-harmony hypothesis of fn. 36 is extended beyond the place of articulation to voicing, irregularities such as PHEC [20] **sazeħ-* ‘three’ → PNHEC **saseħ-* ‘three’ → Had *saso*, Kam *sasu*, Ged and Sid *sase* explicable

[12] A further challenge occurs when one looks at the variants with both **s** and **ʃ** in Burji *saalfad’-*, *ʃaalfad’-* ‘to shy, be feeling fear, become timid’ (Getahun Addo, 2017, p. 735).

as a voicing-harmony effect. This extension goes beyond classic sibilant harmony and is offered tentatively.

The above discussion of PHEC ***ʃ** covers its distribution, inheritance, and contact. Initial **ʃ** has largely merged with **s** except in remnant cases, e.g., PHEC [95] *ʃomb-* 'lung', words without cognates in the rest of EC (except Oromo), e.g., PHEC [523] **ʃam-* 'to rot, to be wet', PNHEC [249] **ʃik' y*, **ʃik' ʔ* 'to approach', PHEC [639] **ʃaf-*, **ʃaf-ʃaf- ʔ* 'to shake', and loanwords, e.g., Amharic *ʃəmbəra* 'chick-peas', *ʃənkurt* 'onion'. There is no intervocalic **ʃ** since it merged with **s**. Exceptions include loanwords such as Amharic *käsäl* 'charcoal' → PNHEC [704] **kaʃil-* 'embers, charcoal'. The irregular **s** and **ʃ** correspondences reflect multiple interacting factors along the contact/inheritance continuum. Inherited distinctions from PHEC underwent both sporadic changes (sibilant harmony, an areal Omotic feature adopted by early HEC speakers) and lexicon-specific influences (borrowing from and contact with Oromo, where PEC ***ʃ** → Oromo **s**).

Finally, for both PHEC ***s** and ***ʃ**, there are no reconstructable geminate consonants, except **ass-* 'to do', which is geminated due to an old fusion with a preceding consonant via derivation from PHEC [425] *ʔ* 'to do' + **-s**. Synchronically, geminate **ss** is found as a reflex of originally single ***s**, at least in the third-person pronouns in Hadiyya and Kambaata, while variation between **ʃʃ** and **ss** is observed in Gedeo and Sidaama in two or three loanwords.

2.1.1.3 Laterals and Trills

The consonant inventory in **Table 1** shows both the lateral **l** and the trill **r**. Their reconstruction to PHEC is secure, supported by consistent retention across all HEC languages except for Hadiyya, which exhibits a unique set of sound changes affecting these sounds.

i. Lateral and trill rules in Hadiyya

   (a) **l** → **r** /**V**_**V**
   (b) **r** → **l** /#_
   (c) **rr** → **ll** (non-initially)

These rules explain the observed reflexes in Hadiyya: **l** → **r** intervocalically, **r** → **l** initially, and geminate **rr** → **ll** non-initially. These Hadiyya-specific changes distinguish it from the other four HEC languages, which preserve the

***l**, ***r** contrast consistently. The changes represent independent innovations within Hadiyya rather than shared NHEC innovations, next to the retention of distinct ***l** and ***r** in Kambaata, Sidaama, and Gedeo cognates in all positions.

PHEC *l

Here are examples of initial and non-initial single PHEC ***l** and its reflexes across HEC:

	[209]		[137]		[705]	
PHEC	***laaf-**	'soft, weak'	***lal-**	'cattle'	***kul-**	'to tell'
Burji	**laafa**	'soft, weak'	**lál-i**	'cattle'	**kulʔud**	'to stutter'
PNHEC	***laaf-**	'soft, weak'	***lal-**	'cattle'	***kul-**	'to tell'
Had	**laafa**	'weary, infirm'	**laro**	'cattle'	**kur-**	'tell'
Kam	**laafa**	'soft, weak'	**lalu**	'cattle'	**kul-**	'tell'
Sid	**laafa**	'weak'	**laló**	'cattle'	**kul-**	'tell, confess'
Ged	**laafa**	'soft, weak'	**lalo**	'cattle'	**kul-**	'tell'

In Hadiyya, examples [137] and [705] show a non-initial reflex **r** of PHEC ***l** in accordance with the rule i(a): ***l** → Hadiyya **r** /**V_V**.

Additional supporting evidence for PHEC ***l**: PHEC [51] **lokk-*, **lekk-* 'foot, leg', [72] **lam-* 'two', [38] **liñ-* 'six', [4] **gulib-, *gilub-* 'knee', PNHEC [33] **osal-ʔ-* 'to laugh', and HEC [136] **eel-* 'pond, pool'.

Geminated ***ll** is retained in Hadiyya.[13]

> PHEC [1] **ill-* 'eye' → Burji *illa*, *ila*; PNHEC **ill-* 'eye' → Had *ille*, Kam *illi-ta*, Sid *ʔille*, Ged *ille*.
>
> PHEC [150] **baall-* 'feather' → Burji *baallee*; PNHEC **baall-* 'feather' → Had *baalleʔe*, Kam *baalli-ta*, Sid *baallicco*, *balle*.
>
> PHEC [276] **ball-* 'brother-in-law, father-in-law' → Burji *bolla*; PNHEC **ball-* 'brother-in-law, father-in-law' → Had, Kam, Sid, Ged *ballo*.

[13] Cognates of [1] and [150] in the rest of EC have a non-geminated **l**, except Oromo *baallee*, *baalʔee* 'feather' and Dhaasanac *baalle*, *baal* 'feather'. The geminated HEC **ll** in [1] is plausibly the reflex of final **l** reduplication. A morphological alternative involving a fossilized suffix **-t** is set aside, as PHEC [175] **kalt-* 'axe, hoe' would be a counterexample.

Other examples of geminated ***ll*** include PNHEC [191] **ull-* 'earth', PHEC [382] **oll-* 'neighbor', PHEC [173] **billaa-t-* 'butterfly', and PHEC [202] **t'all-* 'only, alone'.

l is retained in ***lC*** clusters in Hadiyya.

> PHEC [175] **kalt-* 'axe, hoe' → Burji *kaltee* 'axe'; Had *kalce?e* 'certain part of the plow', Kam *kalta* 'hoe'.
>
> PNHEC [298] **el ʔ* 'to be useful' → Had *e?l-* 'be better, improve', Sid *el-* 'to grant, make donation; use, be useful to ', Ged *el-* 'be better, improve'.
>
> Had [604] *c'alc'al*, Kam *c'alc'al y-* 'to hang'.

In loanwords, non-initial ***l*** is borrowed as **l** in Hadiyya.

> PHEC [349] **gabal-* 'border, limit' → Burji *gabalaa*; PNHEC **gabal-* 'border, limit' → Had, Kam, Sid, Ged *gabala*.
>
> PHEC [605] **galat-* 'thanks' → Burji *galataa*; PNHEC **galat-* 'thanks' → Had, Kam, Sid, Ged *galata*.

The source of [349] is unknown, while [605] is probably from Oromo *galata* 'thanks' given the irregular Hadiyya reflex **l**.[14]

Overall, PHEC ***l*** is retained in initial position, in clusters, and when geminate in all of HEC. It is only changed in intervocalic position in Hadiyya. Thus, it is a secure reconstruction.

PHEC *r

Here are examples of initial and non-initial PHEC ***r*** and its reflexes across HEC:

[14] One might wonder if it was **galt-* hence lack of ***l*** → **r** in Hadiyya? Or was it transferred from Kambaata into Hadiyya?

	[184]		[431]		[110]	
PHEC	***reħ ʔ**	‘to die’	***raʔ-**	‘ripen, be ripe’	***barr-**	‘day’
Burji	**rey-, re-**	‘to die’	**raʔ-**	‘be ripe’	**bari**	‘dawn’
PNHEC	***reħ-**	‘to die’	***raʔ-**	‘ripen, be ripe’	***barr-**	‘day’
Had	**leh-**	‘to die’	**liʔ-is-**	‘become ripe	**balla**	‘day’
Kam	**reh-**	‘to die’	**reʔ-**	‘be ready (food)’	**barraa**	‘day’
Sid	**re-, reʔ-**	‘to die’	**raʔ-**	‘ripe’	**barra**	‘day’
Ged	**rey-, re(ʔ)-**	‘to die’	**raʔ-**	‘ripen, be ripe’	**barra**	‘day’

Rule i(b) states that initial ***r** becomes **l** in Hadiyya, as the above examples demonstrate.[15]

Further examples of non-initial ***r** include: PHEC [189] **k’araar-* ‘be bitter’, PNHEC [42] **k’eer-* ‘long, tall’, PNHEC [127] **kontor-* ‘to snore’, PHEC [122] **gor-* ‘thorn bush’, PHEC [141] **k’ar-* ‘to sharpen’, PNHEC [212] **ar-* ‘husband’, PNHEC [223] **eger* ‘to wait’, PNHEC [224] **t’ur-* ‘dirt of body/clothing’, PNHEC [345] **raar-* ‘shout, scream’, and PNHEC [65] **faras-*, **faras-ʃ-* (sgv) ‘horse’.

In a consonant sequence ***rC**, ***r** does not change in Hadiyya.

> PHEC [85] **arg-* ‘to lend/borrow’ → Burji *arg-*, *erg-* ‘send’; PNHEC **arg-* ‘to send’ → Had *hereg-* ‘to share/to lend’, *arguug-* ‘to chase’, Kam *arg-,* ‘lend/borrow something that can be returned’, *arguug-* ‘chase away’, Sid *arg-* ‘lend/borrowed (other than money)’, Ged *erg-* ‘send’.
>
> PHEC [161] **bark-* ‘head-support, pillow (of wood)’ → Burji *bórk-e*; PNHEC **bark-* ‘head-support, pillow (of wood)’ → Had *barkuma*, Kam *borkaanu*, Sid *barko*, Ged *barko* (cf. Oro *bark-oo* ‘headrest’).
>
> PHEC [220] **c’ark’-* ‘rags, cloth’ → Burji *c’ark’aa* ‘rag, worn-out cloth’; PNHEC **c’ark’-* ‘rags, cloth’ → Had *c’ark’a* ‘cloth, rag’, Kam *c’ark’a* ‘piece of cloth, shred of cloth’, Ged *c’ark’e* ‘rags, cloth’.

Hadiyya [85] *hereg-* ‘to share/to lend’ involves reduplication and insertion of initial **h-** and second vowel **e**. The Hadiyya forms are treated as an inherited

[15] Each of the above examples has one or more related PHEC roots (see the appendices). For set [431], the Hadiyya and Kambaata cognates show a potential vowel ablaut (§2.2.2), and the Sidaama variant *le?-* ‘ripe’ has an irregular initial **l**. Additionally, in PHEC [610] **roob-* ‘hippopotamus’ → Burji *roobee*; PNHEC **roob-* ‘hippopotamus’ → Had and Kam *looba*, Sid *roo?e*, Ged *roop’e*, the Kambaata reflex **l** is irregular and possibly due to Hadiyya influence.

reflex of PHEC *arg-* with internal reshaping (reduplication and **h** prosthesis), not as borrowings. The Oromo cognate *erga* ‘send’ has influenced the semantics and supplied the vowel reflexes in the Gedeo and Burji variants. A similar influence might have been exerted on the vowel reflex (**a** ~ **o**) of the widespread root in the second example [161]. That is in addition to the potential irregular rounding conditioned by the preceding labial. Parallel to PHEC [220] **c’ark’-* ‘rags, cloth’, Amharic has *c’ärk’* ‘cloth’ and Oromo has *c’ark’ii* ‘ragged, cloth’.

Geminated ***rr** becomes Hadiyya **ll** in accordance with rule i(c) ***rr** → **ll** non-initially as seen in PNHEC [110] **barr-* ‘day’ → Hadiyya *balla* ‘day, date, time’ above and the following examples:

> PNHEC [777] **dirr-* ‘to descend, go down’ → Had *dill-*, Kam, Sid, Ged *dirr-*.
>
> PHEC [228] **harr-* ‘donkey’ → Burji *harree*; PNHEC **harr-*, **harr-i-ʃʃ-* ‘donkey’ → Had *halla*, Kam *harri-ta*, Sid *harre*, Ged *hare*.
>
> PHEC [9] **arrab-* ‘tongue’ → Burji *arraaba*; PNHEC **arrab-* ‘tongue’ → Had *allabo*, Kam *arrabi-ta*, Sid *ʔarrabo*, Ged *arrabo* (see PHEC [548] **arraab-* ‘to lick’).
>
> PNHEC [190] **arr-* ‘sun’ → Had *ellin-co*, Kam *arru* ‘sunlight’, Sid, Ged *arijʃo*.
>
> PNHEC [226] **c’aac’urr-*, **c’iic’oorr-* ‘chick(en)’ → Sid *c’aac’c’urre*, Had *c’iic’oolla*, *c’iic’oora*, Kam *c’iic’oorra*.

Based on the cognates and rule i(c), Hadiyya [226] *c’iic’oolla* is the regular variant, while *c’iic’oora* with a single **r** has no parallel. The above examples show that PHEC ***r** changes to **l** (rule i(b))[16] and **ll** (rule i(c)) in Hadiyya in initial position and when geminate, respectively. Otherwise, it is retained across all HEC languages, supporting its secure reconstruction to PHEC.

In conclusion, the changes of the lateral and trill are confined to Hadiyya and can be summarized as follows:

[16] There are two instances of free variation between non-initial **l** and **r** in Hadiyya: PNHEC [397] **kikkil-* ‘armpit’ → Had *kikkiro*, *kilkilo* and PHEC [121] **gol-* ‘wall’ → Had *gorteʔe*, *goratanna*, *goleta* ‘wall’.

		initial	non-initial	_C
***l**	→	**l**	**r**	**l**
***r**	→	**l**	**r**	**r**
***ll**	→	-	**ll**	-
***rr**	→	-	**ll**	-

Single ***l** and ***r** merged with **l** word-initially and **r** non-initially (intervocalically), and geminate ***ll** and ***rr** merged with **ll** (non-initially). Thus, synchronically, Hadiyya has no initial **r** (Tadesse Sibamo, 2015, p. 22) except in a few loanwords (see Ritter, 2007, p. 111).[17] The contrast between the old single ***l** and ***r** is only preserved before some heterogeneous consonants, i.e., in ***lC** and ***rC** clusters in Hadiyya.

2.1.1.4 The Nasal

PHEC *n

The reconstruction of PHEC ***n** is secure despite its defective distribution in initial position. While initial ***n** is rare in native HEC vocabulary, occurring primarily in pronominal forms, e.g., [188] **ni-ʔ-* 'our' and suffixes, the consonant shows consistent reflexes across all five languages in non-initial positions

	[188]		[2]		[54]	
PHEC	***ni-ʔ-**	'our' (pl.poss)	***an-i**	'I'	***min-**	'house'
Burji	**ninka/ninku, ninci/ninta**	'our'	**ʔáṇi**	'I'	**mina**	'house'
PNHEC	***ni-ʔ-**	'our' (pl.poss)	***ani**	'I'	***min-**	'house'
Had	**ni**	'our'	**ʔáṇi**	'I'	**mine**	'house'
Kam	**nii, -nne**	'our'	**án**	'I'	**mini**	'house'
Sid	**ninké, -nke**	'our'	**ʔani**	'I'	**mine**	'house'
Ged	**noʔoki, noʔot'i**	'our'	**ani**	'I'	**mine**	'house'

Moreover, broader East Cushitic comparanda demonstrate regular correspondences for ***n** (see PEC [54] **min-* 'house'), confirming its reconstructibility to PHEC. The restricted distribution in initial position likely reflects a lexical gap rather than a phonotactic constraint, as loanwords freely allow initial **n**, e.g., Oromo [152] *naacca* 'crocodile' in HEC.[18] The nasal ***n**

[17] The evidence for reconstructed initial ***r** is limited in the available data.
[18] It is also found on the isolated roots PNHEC [1032] **nigga* 'tendon' and PHEC [1033] **naac'a* 'joke'.

also plays a productive role in HEC morphology, e.g., nominal marking §4.1 (see also genitive §4.1.4.1 and epenthetic **-n-** §3.4.2), further supporting its secure reconstruction.

Other cognate sets show: PHEC [745] **-n* 'with', PHEC [66] **anuun-* 'breast', PHEC [67] **ang-* 'hand, arm', PHEC [5] **san-* 'nose', and PHEC [628] **ann-* 'father'.

2.1.1.5 Summary

The reconstruction of alveolar and post-alveolar consonants in PHEC provides insights into the phonological history of the Highland East Cushitic languages. While the stops **t** and **d** exhibit clear and consistent reflexes, the fricatives, particularly **z** and **ʃ**, present a more complex picture of merger, retention, and potential borrowing. The systematic sound changes affecting lateral **l** and trill **r** in Hadiyya highlight the independent phonological innovations within that language. The nasal **n** has a limited distribution word-initially, yet remains stable in other positions and even plays a significant role in morphology. The following section (2.1.2) examines the non-pulmonic alveolars and post-alveolar, where contact effects and analogical developments require particularly careful analysis.

2.1.2 Non-pulmonic Alveolars and Post-alveolar

Having established the pulmonic consonants, this section examines the non-pulmonic alveolar and post-alveolar consonants, where the relationship between inheritance and contact is more complex. The existing literature describes them as glottalized, glottalic, ejective, retroflex, and implosive consonants, but here I use the term "non-pulmonic" consonants (including the velar **k'**). This reserves "glottalized" for secondary non-pulmonics derived through morpheme boundary alternation processes (detailed in §1.5.2). This section begins with a synthesis of key discussions in the existing literature on the non-pulmonic consonants, then presents an analysis of the historical development of these consonants in the HEC languages.

Previous scholars have attempted to reconstruct non-pulmonic consonants across various Cushitic and Omotic proto-languages. The following list synthesizes those key reconstructions:

Proto-Language	Non-pulmonics	Source
Proto-Highland EC	***d', *t', *c', *k'**	Hudson (1989)
Proto-Lowland EC	***D, *q**	Black (1974, p. 280)
Proto-East Cushitic	***d', *d'$_1$, *k'**	Sasse (1979, p. 5)
Proto-East Cushitic	***b', *d', *t', *c', *j', *k', *g'**	Ehret (1991, p. 273)
Proto-East Cushitic	***ɗ, *D**	Bender (2020)
Proto-Cushitic	**-**	Bender (2020)
Proto-Cushitic	***p', *t', *ts', *c', *k', *k^w'**	Ehret (1987, p. 163)
Proto-Omotic	***p', *ɓ, *t', *ts', *c', *k'**	Bender (1988, p. 139)
Proto-Omotic	***p', *d', *ts', *s', *c', *k'**	Ehret (1995, pp. 10–12)
Proto-South Omotic	***p'/*b', *t'/*d', *s'/*ts', *c', *k'/*g'**	Fleming (1988, pp. 166–172)

Hudson (1989, p. 9) provides phonetic descriptions of the non-pulmonic consonants attested in HEC languages: ejectives (**p'**, **t'**, **c'**, **k'**); glottalized and somewhat ingressive **d'** in Burji, with **b'** found only in Burji; glottalized and retroflex **d'** in Gedeo and Sidaama; and dialectal variants of glottalized sonorants (**'l ~ ʔl**, **'r ~ ʔr**, **'n ~ ʔn**) in Burji, and possibly **'l ~ ʔl** in Gedeo. These phonetic distinctions become relevant when evaluating whether **d'** reflexes represent inherited phonemic contrasts or contact-induced phonetic variation (see discussion under PHEC *t').

Bender (2020, pp. 155, 138, 176) has not reconstructed any non-pulmonic consonants for Proto-Cushitic and for Proto-East Cushitic only an implosive *__ɗ__ and, provisionally, a retroflex ***D**. Black (1974, p. 280) similarly reconstructed only a retroflex *__D__ and a uvular *__q__ for Proto-Lowland East Cushitic.

Early studies have used the distribution of non-pulmonic consonants to classify typological features within Cushitic languages. According to Fleming, as cited in Hudson (1976, p. 246), glottalized ejectives distinguish HEC from Lowland East Cushitic (LEC), the latter being characterized by pharyngeal consonants. This feature groups HEC with Oromo-Konsoid-Werizoid rather than other Lowland languages. Fleming (1976, pp. 306–307) attributes this HEC vs. LEC distinction to Omotic influence on HEC. The distinction between HEC and LEC is largely attributed to the influence of Omotic languages on HEC.

Fleming's claim has been the subject of an ongoing debate about whether the non-pulmonic consonants in Highland East Cushitic languages are inherited

or borrowed.[19] Arvanites (1991) argues for an inherited status of **t'** and **c'** in addition to **d'** and **k'** from Proto-East Cushitic. She rejects Fleming's hypothesis of large-scale borrowing from Omotic into HEC. Sasse (1979, cited in Arvanites, 1991, p. 122) supports the inheritance hypothesis over borrowing when good correspondences can be found in EC.

Crass (2002) also argues for the hypothesis of contact. He suggests that the non-pulmonic consonants disappeared over time in many languages and were re-established later due to language contact. Crass attributes the re-establishment of the non-pulmonics in various Cushitic languages to contact with Amharic, Tigrinya, or Oromo (Crass, 2002, pp. 1686, 1688). Although claims in Crass (2002) lack evidence, they are worthy of further investigation.

The above overview reveals the historical complexity of non-pulmonic consonants in (Highland) East Cushitic languages. Scholarly approaches to their reconstruction vary considerably. That is because, unlike most of the pulmonic consonants examined in §2.1.1, the non-pulmonic consonants show substantial irregular reflexes, making the contact-inheritance analysis particularly critical for this consonant class. The present analysis applies the contact/inheritance continuum articulated in §1.5.4, which recognizes inherited proto-phonemes and contact-influenced developments as endpoints on a continuum rather than mutually exclusive categories, to be evaluated case-by-case based on distributional evidence.

Following this approach, the analysis examines each proto-phoneme through multiple lenses simultaneously: (1) regular correspondences across HEC languages backed up with wider EC context to establish inherited phonemes, (2) distributional irregularities to flag potential contact influence such as geographic clustering or irregular reflexes, and (3) areal context from external comparison to Oromo, Omotic, and other neighboring families (see Hudson, 1989, p. 8). Where correspondences are regular and evidence substantial, the pattern is treated as inherited per the parsimony principle presented in (§1.5.4), while irregular reflexes and suspicious distributions reveal contact-influenced sound changes.

Applying the probability scale (§1.5.1) and contact/inheritance continuum (§1.5.4), the evidence supports reconstructing two PHEC proto-phonemes in

[19] According to Fleming (1976, p. 307) "[p]honologically... adjacent Semitic, Burji-Sidamo [HEC], and Omotic languages share highly similar sets of features, especially four or five glottalic ejectives".

the non-pulmonic alveolar-post-alveolar series: ***t'** and ***c'** (reasonable). Both show regular correspondences across HEC languages, substantial internal supporting evidence, and promising East Cushitic parallels. The analysis examines **d'**, attested in Burji, Gedeo, and Sidaama, and analyzes it as a contact-influenced development within the **t'** correspondence set.

PHEC *t'

The reconstruction of PHEC ***t'** is reasonable and supported by regular reflexes across all five HEC languages, initial and non-initial distribution patterns, and clear East Cushitic cognates establishing ***t'** as a potential PEC proto-phoneme. However, ***t'** exhibits remarkable reflex diversity in Burji, Gedeo, and Sidaama, where irregular **d'** and **c'** reflexes arise from intensive contact with neighboring Oromo, Omotic, and Amharic languages. These irregular correspondences represent contact-induced developments in the inherited lexicon, illustrating the contact/inheritance continuum framework established in §1.5.4.

The following analysis presents evidence sequentially: regular correspondences across different positions in a word establishing the proto-form, comparative East Cushitic evidence demonstrating inheritance, and a detailed examination of irregular reflexes with explicit contact explanations. This integrated presentation reflects the reality that the same lexical sources can exhibit both regular (inherited) and irregular (contact-influenced) reflexes across the language family.

Arvanites (1991, p. 136) extracts the non-pulmonic PEC consonant ***t'** (and ***c'**) from the reflexes of Sasse's problematic PEC ***d'**$_1$. According to the available data (§1.3), **t'** occurs in approximately 10 instances in Burji, while **c'** appears in over 25. Excluding Burji, there is additional evidence of **t'** across different word positions in NHEC. Below is a summary of the attested series of correspondences, their positions, and frequency covered in the analysis:

	Position	Burji	Had	Kam	Sid	Ged	Frequency
1.	Initial, non-initial	**t'**	**t'**	**t'**	**t'**	**t'**	Common
2.	Initial	**d'**	**t'**	**t'**	**d'/t'**	**d'**	Uncommon
3.	Initial	**d'**	**ø**	**ø**	**ø**	**ø**	Uncommon
4.	Non-initial	**ʔ/d'**	**ø/ʔ**	**ø/ʔ**	**ø/ʔ/d'**	**ø/ʔ/d'**	Common (suffix)
5.	Initial	**c'**	**t'**	**t'**	**t'**	**t'**	Uncommon
6.	Initial	**c'**	**t'**	**t'**	**c'**	**c'**	Uncommon

Except for the first series, all include irregular reflexes. The slash between variants, e.g., Sidaama's initial **d'**/**t'**, indicates that both are attested within that series in different lexemes or even the same lexeme in the case of Sidaama. This analysis will follow the above order: starting with series 1 (with regular **t'**), then series 2–4 (with irregular **d'**), and finally series 5–6 (with irregular **c'**).

Before turning to the mixed **c'** series, I state the diagnostic used here. I reconstruct inherited PHEC ***c'*** only where **c'** is the regular reflex across all five HEC languages; those sets are treated separately below (see PHEC *c'). Where Hadiyya and Kambaata retain **t'**, and **c'** is confined to Burji or the southern languages (Gedeo and Sidaama), I treat **c'** as an irregular development within the ***t'*** correspondence set. I invoke contact when mixed HEC correspondence, a restricted southern distribution, the lack of EC parallels, and a plausible Oromo or Omotic model coincide.

Regular reflexes of PHEC *t'

The following examples provide evidence for the reconstruction of ***t'*** in initial position with regular correspondences across HEC languages.

	[351]		[308]		[269]	
PHEC	***t'uut'-***	'to suck'	***t'uur-***	'to milk'	***t'aʔ-, t'eʔ-***	'to extinguish' 'to diminish'
Burji	**t'uunt'-**	'absorb'	**t'uur-**	'to milk'	**t'ay-, t'ey-**	'extinguish' 'dry out'
PNHEC	***t'uut'-***	'to suck'	**-**	-	***t'aʔ-, t'eʔ-***	'to extinguish' 'to diminish'
Had	**t'uut'-**	'to suck'	**-**	-	**t'aʔa**	'be depleted/ diminish'
Kam	**t'uut'-**	'to suck'	**t'uur-**	'to milk'	**t'aʔ-, t'eʔ-**	'extinguish' 'be depleted'
Sid	**t'uut'-**	'to suck'	**t'uur-**	'to milk'	**t'eʔ-**	'be depleted'
Ged	**t'uut'-**	'to suck'	**-**	-	**t'eʔy-**	'be incomplete'

Set [351] presents **t'uut'-* 'to suck' reconstructed from regular correspondences in HEC languages. Certain sound changes are notable within these sets. For instance, in [351], the extra **-n-** in Burji *t'uunt'-* is an inserted sound (§3.4.3). The common vowel variation before the glottal stop in [269] is discussed in §2.2.2.1.

In non-initial position, evidence for PHEC ***t'*** from all HEC languages comes from the following reconstructions:

PHEC [304] **wit'-* 'scatter' → Burji *wat'ar-*, *wot'ar-*; PNHEC **wit'-* 'to sow' → Had *wit'-*, Kam *wit'-*, Sid *wit'-*, Ged (*wiisa*).

PHEC [240] **k'ort'umm-* 'fish' → Burji *k'urt'ummee*; PNHEC **k'ort'umm-* → Had *k'urt'umeʔe*, Kam *k'urt'ummeema-ta* / *k'urt'ummee-ta*, Sid *k'ult'uʔme*, Ged *k'ult'uʔme*.

PHEC [10] **hat't'iʃ y* 'to sneeze' → Burji *hat't'iʃee-d'-* (cf. *ha**tt**iʃeed'-*, *ha**tt**iʃeys-* (Roba Dame & Wedekind, 2008, p. 37)), *hat'iʃ-*; PNHEC **hat't'iʃ y* → Had *hat'iʃʃuy-*, *hat't'eʃ-*, Kam *hant'iʃʃ-aʔ*, Sid *hant'iʃʃ-iʔr-*, *hant'iʃʃo*.

In set [304], the Gedeo form *wiisa* may contain a causative **s**, with double **ii** possibly indicating the reduction of **t's** clusters. The onomatopoetic nature of set [10] might have contributed to irregularity in the quantity and quality of some of the correspondences.

Additional initial and non-initial evidence of **t'** appears within NHEC in examples such as PNHEC [55] **t'ut'-* 'black ant', [744] **fit't'-* 'to disperse', [52] **hit'-* 'grass (sp.)', [613] **t'een-* 'rain', [661] **t'umm-* 'peace', [730] **t'aʔm-* 'to ask, visit', [224] **t'ur-* 'dirt of body/clothing', and [75] **t'eʔm-* 'to taste good'.

The regularity of correspondences across initial and non-initial examples supports the reconstruction of PHEC ***t'**, which is further supported by some EC parallels with a corresponding **t'**, albeit with some irregularities in the corresponding sounds other than **t'**:

PHEC [222] **t'ok-* 'to burst': Gaw *t'uh-* 'to sprout', Dah *t'oʕ-* 'to pain/cut'.

PHEC [10] **hat't'iʃ y* 'to sneeze': Arb *het't'iris-*, Gaw *ħat't'is-*, Bay *hant'irʃe* 'sneeze'.

PNHEC [55] **t'ut'-* 'black ant': Yaaku *t'oroʔtɛ* 'driver or safari ant', Dah *t'at'e* 'small ant', Oro *t'iint'-* 'maggot'.

PHEC [250] **t'iib-* 'to push': Oro *t'ibiira* 'twist (like in making ropes)', Gaw *t'iip-* 'to squeeze', Dah *t̪'uuʕ* 'to press/throttle', *t'uub* 'squeeze'.

Beyond HEC, according to Ehret (1987, p. 34), Yaaku **t'** and Oromo **d'** are reflexes of PC ***t',** while Oromo words with **t'** are possibly borrowings.[20]

Certain HEC examples exhibit one parallel in a neighboring language, such as PHEC [351] **t'uut'-* 'to suck', with a comparable stem in Oromo *t'uut'a* 'suck', and HEC [269] **t'aʔ-* 'to extinguish', maybe with Gawwada *ɗoʔ-* 'to dry up and crack'. PHEC [304] **wit'-* 'to sow' might have been transferred into Bayso as *wit'-* 'sow, seed' and it is comparable to Hamar *yiit'* 'sow, planting seeds by scattering them on the earth' (Petrollino, 2016, p. 318) and Sheko *yiits'* 'to sprinkle' (Hellenthal, 2010, p. 975). These HEC-only parallels point to external (non-HEC) origins or HEC innovations transferred into the neighboring languages.

Six of the reconstructed roots with PHEC ***t'** have no comparable parallels in EC or the neighboring Omotic/Amharic languages: PHEC [308] **t'uur-* 'to milk',[21] [744] **firt'-* 'to disperse',[22] and PNHEC [52] **hit'-* 'grass (sp.)', PNHEC [613] **t'een-* 'rain', [661] **t'umm-* 'peace', and [730] **t'aʔm-* 'to ask, visit'. In these cases, the absence of parallels suggests that some HEC roots may be innovations or borrowings from an unknown source.

A few of the above HEC roots have parallel examples with a corresponding **t'** in the neighboring languages, particularly in Oromo:

HEC				**Parallels**		
Gedeo	[10]	**het't'if-at-**	'sneeze'	Oromo	**hat't'iffatu**	'sneeze'
PHEC	[240]	***k'ort'umm-**	'fish'	Oromo	**k'urt'ummii**	'fish'
PHEC	[351]	***t'uut'-**	'to suck'	Oromo	**t'uut'a**	'suck'
				Amharic	**t'ut**	'breast'
PHEC	[269]	***t'aʔ-**	'extinguish'	Wolaytta	**t'ay-**	'to get lost'

In [10], Gedeo *het't'if-at-* exhibits a typical Oromo sound change, ***s** → **f**, which might be a result of later contact influence. PHEC [240] **k'ort'umm-* 'fish' is another Oromo loan from *k'urt'ummii* 'fish' (Sasse, 1982, p. 129; see Hudson,

[20] Sasse (1979, pp. 29–30) proposes a medial PHEC ***t'** but for different reflexes, namely Hadiyya **ʔ**, Sidaama and Gedeo **d'**. These reflexes rather include various suffixes discussed in Ch. 4.

[21] PHEC [308] **t'uur-* 'to milk' might be comparable with Burunge *ɬuruʔud* and Alagwa *ɬuruʔut* 'sip, suck' in the West-Rift Cushitic languages (Kießling & Mous, 2003, p. 275).

[22] PHEC [744] **firt'-* 'to disperse', [721] **fir-* 'relative', and [162] **fic'c'-* 'to scatter' might be related. Kießling & Mous (2003, p. 108) reconstruct PWR **fiits'-* 'sweep, level', which they link to Sidaama **fit'* 'comb' and Somali *fid'* 'comb'. Those are rather linked to PNHEC **fil-ʔ-* 'comb' in this study. Lamberti & Sottile (1997, p. 357) interpret Zaysse *pits'-* 'sweep' to include the benefactive suffix ***-ʔ-** that has caused the glottalization of the dental in "the old West Cushitic stem" **bit-*, after which ***-t'-** became **-ts'-**.

1989, p. 8). However, it might be a very recent or a very old loan because it is found all over HEC. PHEC [351] **t'uut'-* 'to suck' may be an original PHEC root, but it is also found in Oromo *t'uut'a* 'suck' and Amharic *t'ut* 'breast' (see Burji *t'uunt'-* 'to suck' and Zaysse *ts'unts'-* 'suck out' (Sasse, 1982, p. 183)). The same applies to PHEC [269] **t'aʔ-* 'to extinguish' with its parallels in Wolaytta *t'ay-* 'to get lost' and other Ometo languages, raising questions about the direction of transfer. Since the latter two are regularly reconstructed for PHEC, they must be inherited PHEC roots.

Irregular d' reflexes of PHEC *t'

The preceding analysis establishes the reconstruction of PHEC ***t'** based on regular correspondences in different word positions and East Cushitic cognates. However, Burji, Gedeo, and Sidaama exhibit irregular **d'** reflexes in specific lexical sets, requiring detailed examination.[23] Such irregular patterns correlate with documented Oromo contact and do not represent a separate proto-phoneme.

An example of the irregularity is that Burji initial **d'** may correspond to either **d'** or **t'** in Sidaama, as illustrated in the following examples:

[23] The synchronic description of **d'** ranges between implosive and sometimes a retroflex in Gedeo, Sidaama, and Burji (Hudson, 1976, p. 248; Bender, 2020, pp. 116–117; Treis, 2008, p. 27). Yri & Pepper (2019) state, in Sidaama, the "…implosive can be described as having an oral and a glottal gesture [glottalized retroflex stop (Hudson, 1976, p. 248)]. Following a liquid or a nasal, it undergoes (underwent) reduction, i.e., loss of its oral gesture and metathesis, probably in that order. The result was a cluster of glottal + liquid/nasal". This reduction hypothesis is rejected as explained in §§4.2.3, 3.3.2.

	[214]		[250]		[128]	
PHEC	***t'aban ʔ-s**	'to slap'	***t'iib**	'to push'	***t'eeb ʔ**	'be thirsty'
Burji	**d'abanaas**	'to slap'	**d'iib-**	'to push'	**d'aybood'-**	'be thirsty'
PNHEC	***t'aban ʔ-s-**	'to slap'	***t'iib-**	'to push'	***t'eeb ʔ**	'be thirsty'
Had	**t'aban-s-**	'to slap'	**t'iib-**	'to push'	**t'eebeʔ-**	'be thirsty'
Kam	**t'aban-s-**	'to slap'	**t'iib-**	'press'	**t'eebeʔ-**	'be very thirsty
Sid	**d'aʔa assa**	'to slap'	**t'iibi**	'to push'	**-**	-
Ged	**d'abanaas**	'to slap'	**d'iib-**	'push'	**d'eebota**	'thirst'
cf. Oro	**d'aana**	'beat on'	**d'iib-**	'push'	**d'eebotee**	'thirsty'

Examples [214], [250], and [128] are the only examples supporting the putative initial PHEC ***d'***, though the Oromo parallels reveal the source of the irregular **d'** reflexes in Burji, Gedeo, and Sidaama. Set [214] provides the strongest evidence in favor of a PHEC ***d'*** because it has the reflex **d'** in three HEC languages. The other two examples [250] and [128] lack a supporting **d'** reflex in Sidaama. Hence, they are all reconstructed to P(N)HEC ***t'***, analyzing the **d'** reflex as irregular, attributed to Oromo influence on the three languages and supported by further evidence below.

Overall, the sound **d'** is predominantly found in the southernmost language, Burji. It appears less frequently in Gedeo and is least common in Sidaama. In Sidaama, initial **d'** is sometimes in free variation with **t'** as in the examples below.[24]

	t'		d'	Gloss
[417]	**t'ibba**	~	**d'ibba**	'disease'
[549]	**t'aam-**	~	**d'aam-**	'to give a message to'
[201]	**t'agge**	~	**d'agge**	'story, miracle'
[734]	**t'ara**	~	**d'ara**	'false'
[159]	**t'aawa**	~	**d'aawa**	'to erect, establish'

Wedekind (1980, p. 149) identifies the potential sources of **t'** ~ **d'** variation by noting that [201] *t'agge* is from Northern Sidaama, while *d'aggi* 'news' is from Southern Sidaama varieties.[25] Thus, the Southern Sidaama forms with a

[24] Hudson (1976, p. 248) describes **d'** as being primarily glottalized in both Gedeo and Burji, somewhat less so in Sidaama, and entirely absent from Hadiyya and Kambaata. This description of **d'** aligns with its distribution from south to north.
[25] Note that the above [201] examples are from a different source, hence the slight variation in transcription.

corresponding **d'** are the most influenced by Oromo/Burji (see Sasse (1979, p. 30)). Further south, Oromo contact with Gedeo is evident in plenty of loanwords, e.g., [346] Gedeo *d'uufo* 'fart' (← **d'uus-*) since ***s** → **f** is an Oromo innovation. Further examples with an irregular initial **d'** attested in Burji, Gedeo, and Sidaama only are found in [549] (Oromo loan), [734], and [78].

In a rare secondary series, Burji has **d'**, while other HEC languages display the reflexes **ø**/**ʔ** in initial position. This series appears in only two or three comparanda, suggesting a potential fossilized reflexive prefix (§4.2.3).

Sasse (1982, p. 62) reiterates Hudson's (1976, p. 272) connection of the verb *d'ayb-ad'-* 'be warm' to the common EC stem *iib-* 'be warm' in Burji and interprets the initial **d'** as a fossilized middle voice prefix, similar to that in *d'irk'-* 'give back'. Hayward (1984, p. 84) confirms that "the only languages having prefixed root extensions are Saho, ᶜAfar and Burji… when due consideration has been given to morphophonological alternations the ᶜAfar and Saho forms can only be accounted for by means of an underlying prefix with **t**, while the Burji forms… clearly contain a prefix with a **ɗ**....". This phenomenon has been observed in examples like:

> PHEC [298] **al ʔ* 'to be useful' → Burji *d'alʔ-* 'to be useful'; PNHEC **el ʔ* → Had *eʔl-* 'be better, improve', Sid *el-* 'to grant, make donation, use, be useful to', Ged *el-* 'be better, improve'.
>
> PHEC [306] **y ebb-* 'be(come) hot/warm' → Burji *d'aybad'*; PNHEC **y ebb-* → Had, Kam, Sid *iibb-*, Ged *eebb-*.

The above evidence, in addition to Burji *d'irk'*, (intr) and *rik'k'-* 'return' (tr), supports the observation that there are reflexives once formed with a prefix in Burji (see Sasse, 1982, p. 65). Further reiteration of Hudson's observation is found in Hetzron (1980, p. 95), who states that "[n]ormally in Cushitic derivational affixes are postverbal. Yet in Saho and Afar, prefix-conjugated verbs have prefixed derivational affixes… G. Hudson (1976:272) has shown that even Burji ... has some vestiges of original derivational prefixes, now lexicalized, e.g. ḍirk' 'return (intr)', vs. rikk' 'return (tr.)' with a reflexive prefix (which is now a suffix -(a)ḍ) ." Hetzron provides further evidence of fossilized prefixes "[i]n Rendille, the verb -tVḥ 'become' may have been formed as a reflexive (medium) of -daḥ 'say' ... and the root -sil(l) 'lay down' could be a causative of -al(l)- 'be placed, exist, remain'". The reconstructed PHEC reflexive is ***-ʔ** (§4.2.3.1), which means Burji **d'** in the prefix is still unique and irregular within HEC, suggesting Oromo as the source (§4.2.3.4). An

alternative potential source for the NHEC reflexes could also be the derivational verbalizer/support verb ***ʔ** since [306] **y ebb-* 'be(come) hot/warm' above also includes the other verbalizer/support verb ***y** (§§4.2.4, 4.2.5).

The above evidence signifies the distinction between Burji and NHEC based on both C_1 and V_1. Additionally, it raises questions such as how many other cases there are of such prefixes beyond Burji[26] and how much that contributes to the idiosyncratic correspondences.

Looking at the reflexes in non-initial position, HEC languages demonstrate an even more complex set of reflexes with irregular **d'**, which reconstruct to PNHEC and PHEC ***ʔ**. Such analysis is contrary to Bender (2020, pp. 116–117) and Hudson (1989, p. 7), who reconstruct a PHEC ***d'** based on its presence in Burji, Gedeo, and Sidaama, with Hadiyya and Kambaata showing **ʔ** as a reflex. Those reflexes are found mainly in a series of non-initial correspondences with irregular reflexes, as in the following examples:

	[348]		[168]	
PHEC	***c'iiʔ-**	'bird'	***heeʔ-**	'to exist, to live'
Burji	**c'iid'aa**	'bird'	**(heʔ-, yed'-)**	'here it is!, exist'
PNHEC	***c'iiʔ-i-ʃʃ-**	'bird'	***heeʔ-**	'to exist, to live'
Had	**c'iiʔ-cco**	'bird'	**heeʔ-**	'live, exist'
Kam	**c'iiccu-**	'bird'	**heʔ-**	'live'
Sid	**c'eʔicco**	'bird'[27]	**heed'-**	'exist, be'
Ged	**c'iʔa**	'bird'	**hed'-**	'exist, be there'

In comparanda [348], **d'** occurs in Burji only, whereas in [168] it is found in Sidaama and Gedeo examples (the Burji cognate is doubtful). Sasse (1979, pp. 29–30) proposes a medial (non-initial) PHEC ***t'** for the correspondence set of Hadiyya **ʔ**, Sidaama, and Gedeo **d'**, which resembles the correspondences in [168]. However, the glottal stop in [168] involves retained and attached reflexive/derivational glottal stops, while the **d'** is a reflexive/derivational suffix in Sidaama and Gedeo (§§4.2.3). In [348], Burji *c'iid'aa* 'bird' is best treated as an isolated Burji irregular non-initial reflex of ***ʔ** rather than as evidence for reconstructing medial PHEC ***t'** or ***d'**.

[26] A comparable instance is found in Oromo [43] *d'uga* 'drink' vs. PS **uk-* 'to drink'.
[27] Sidaama *c'eʔicco* 'bird' is contractable into *c'ïcco* (Gasparini, 1983, p. 56).

There are many more examples of non-initial **d'** found on derived verbs; thus, they are independently added suffixes (§§4.2.4, 4.2.5) as the following correspondences demonstrate:

[256]	**ø**	**-ʔ**	**-t**	**-d'**	**Gloss**
PHEC	**-**	***ham ʔ**			'to slander'
Burji				**ham-ad'-**	'slander'
Had	**ham-**				'slander'
Kam		**heʔm-**			'slander, gossip'
Sid	**heme**	**heʔm-**			'backbiting', 'slander'
Ged			**ham-at-**		'slander'
cf. Oro			**hammaataa**	**hamad'd'a**	'bad off', 'slander'

[450]	**ø**	**-ʔ**	**-t**	**-d'**	**Gloss**
PHEC	***ʃaal-**				'be shy, be ashamed'
Burji		**ʃa(a)lʔ-**		**saalfad'-**	'be shy', 'be embarrassed'
Kam	**saal-**				'be shy, ashamed'
Sid	**saal-**		**saalfat-**		'be shy, ashamed'
Ged			**saalfat-**		'be shy, ashamed'
cf. Oro				**saalf-ad'd'a**	'be shy, ashamed'

[1030]	**ø**	**-t**	**-d'**	**Gloss**
PHEC	***ħuk'-**			'be(come) thin'
Burji			**huk'k'ad'-**	'to get thin'
Ged		**huk'k'ata**		'be slim'
cf. Oro		**huk'k'at-**	**huk'k'ad'd'a**	'thin', 'become thin'

The above examples show that the correspondence sets include at least **ø**, **ʔ**, **t**, and **d'**. Again, Oromo influence is noticeable where Burji has **d'** and Gedeo and Sidaama have a corresponding **t** both of which are suffixes given the formations provided in the Oromo dictionaries (Gragg, 1982; Tilahun Gamta, 2004; Stroomer, 2001).[28] The other HEC languages have corresponding

[28] Oromo **-t-** could be interpreted as a variant of the middle voice verb extensions of Borana (Marsabit) and Waata dialects reported by Stroomer (1987, pp. 153, 155) that produce variants **-ad'd'-** (for first-person singular) and **-at-** (elsewhere) with occasional variation in vowel quality. It might also be a different derivational suffix like the one found in HEC (§4.2.5.3).

underived stems with **ø**, such as in Hadiyya, Kambaata, and Sidaama, and stems with the reflexive/derivational **-ʔ** in Hadiyya and Kambaata ([450]).

While Burji **-d'** is a loan from Oromo, Sidaama, and Gedeo **-t** is likely a contact-induced retention of an agentive denominative (cf. §4.2.5.3). They were likely already cognates in Gedeo and Sidaama and have been influenced by the Oromo words ending in **-t**.

See also [497], [177], [340], and [698] for further examples.

The irregular correspondences in initial and non-initial positions, the phonological processes at morpheme boundaries (§3.3.1.5), and the influence of Oromo on certain Sidaama and Gedeo forms and loans in Burji challenge any proposal of a consistent reconstruction for ***d'** for PHEC.

The wider East Cushitic context reveals greater complexity but could be synthesized as the most common correspondences of PHEC ***t'** are probably retained **t'** in Southern East Cushitic (Yaaku-Dullay), and innovated **d'** in the Lowland East Cushitic branch. Since the EC parallels with a corresponding **t'** have been discussed above, the following paragraphs will be dedicated to the ones in which the regular and irregular reflexes of PHEC ***t'** correspond to **d'** (realized as **d'**, **ɗ**, or **ɖ**) in the rest of EC.

> PHEC [222] **t'ok-* 'to burst': Oro *d'oʔa* 'burst, explode', Dir *ɗeh* 'grow', Mos *he-ɗehe* 'sprout', Kon *ɗeh-* 'sprout', Afar *ɗokʕ-* 'bruise, dent', Arb *d'oʔ-* 'burst'.
>
> PNHEC [55] **t'ut'-* and PHEC **t'uut'-* 'black ant': Dir *ɗuuɗɗéṭ* 'small black ants, black beads', Mos *ɗuɗe-ta* 'ant', Saho *ɖuuɖe* 'ant', Afar *ɗuuné* 'ants', Ren *ɖúuɖe* 'black ant(s)', Som *ɖuurcaso* 'ant species reddish orange in color', *ɖuuɖi* 'a kind of louse', Maay *ɖuuri* 'ants', Boni Kilii *d'óói* 'ant sp. (small)'.
>
> PNHEC [75] **t'eʔm-* 'to taste good': Oro *d'amd'ama*, 'taste lightly', *d'an-d'am-* 'to taste', Saho *ɖaʕame* 'to taste', Afar *ɗaʕame* 'to taste', Bay *t'am-*, Som *ɖaɖami* 'taste', Maay *ɖeɖaŋ* 'taste', Boni *d'id'imiy-* 'taste', Das *d'ád'am* 'taste'.
>
> PHEC [214] **t'aban ʔ*, **t'aban ʔ-s-* 'to slap' ← PHEC [106] **t'aban-* 'side, cheek': Kon *ɗapn-a* 'side (of face, body, mountain, river)', Oro *d'amn-a* 'jaw', Saho *ɖaban* 'ceek, side', Afar *ɗában* 'cheek, side', Ren *ɖában*

> ‘side’, Som *ɋaban* ‘cheek’, Maay *ɋabaŋk* ‘cheek’, Boni *d’áŋ* ‘side, direction’.
>
> PHEC [250] **t’iib-* ‘to push’: Kon *ɗiip-p-* ‘grasp, squeeze’, Som *ɋiiɋiib-* ‘bring thighs together to close crutch [crotch]’, Saho *ɋibɽo* ‘catching, hold’, Afar *ɗibbaanise* ‘connect together’, PEC **d’ˌiib-* ‘squeeze, press’ (Sasse, 1979, p. 29).

In [222], the reflexes in Sidaama and Gedeo show irregular **d’** in comparison to the rest of HEC.

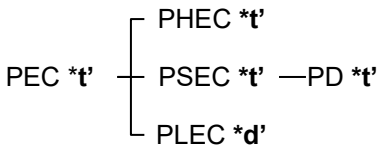
Some of the above examples have variants in both HEC and wider EC recorded with a consonant other than **d’** such as **c’**, **j’**, **ʃ**, **c**, **t**, and **d**.

> [222] Burji *tokkaʔ-* ‘to leak/drop’, *dar-ʔ-* ‘burst’, Sid *taʔi* ‘burst, crack, explode’, Ged *doʔ-* ‘burst’.
>
> [55] Burji *c’uuc’c’ee* ‘black ant’ and Had *c’uuma* ‘big black ants’, Boni Baddey *j’úúrə*, Boni Bireeri *c’úúrə* ‘ant sp. (small)’, Bayso *tunce* ‘red ant’.
>
> [75] Sid *c’oommi* ‘taste good’, Boni *iʔimiy*, Afar *taam-* ‘to taste’, Ren *camcama* ‘(to) taste’, Das *ʃam* (adj) ‘taste, strong taste’, Dah *tem-* ‘to taste’.
>
> [250] Dir *c’iip-* ‘squeeze’, Mos *c’iib-eeca* ‘squeeze, wring out’, Saho *dufuwe* ‘to push’, Afar *duuf-* ‘forceful push’.

Because these forms coexist with **d’** reflexes in the same language, I do not treat them as evidence for a single regular East Cushitic correspondence. At most, they indicate mixed dialectal variants or secondary lexical interference. In [55] PHEC **t’uut’-* ‘black ant’, Burji and Hadiyya **c’** diverge from reconstructed ***t’**; however, Boni parallels also have **c’**. The variation in the reflexes of this set might be due to references to different species of ants, as the glossing suggests, or multiple contact sources. The second most common irregular EC correspondence of PHEC ***t’** is **c’**, such as in [55], and will be addressed in detail in the next series of irregular correspondences.

In Lowland EC, the most common sound is **d’**, which corresponds to **t’** in HEC, SEC, and Dahalo languages. Hence, PHEC ***t’**, PSEC ***t’**, and PLEC ***d’**

reflexes support the reconstruction of PEC ***t'*** (see PC ***t'*** in Ehret (1987, p. 34)) according to the traditional subgrouping:



In conclusion, **d'** is a phoneme in Burji, Gedeo, and Sidaama; however, it is not reconstructable to P(N)HEC ***d'*** due to the irregular correspondences and evident influence of Oromo over these three languages. Thus, initial **d'** in Burji, Gedeo, and Sidaama (i.e., from series 2–3 (see *Irregular **d'** reflexes of PHEC **t'***)) results from independent contact with Oromo, while the correspondences in non-initial position from series 4 (i.e., the reflexive/derivational suffix) developed from contact-induced change and via a regular fusion process of **r** and **ʔ** in Sidaama and Gedeo, as discussed in the next chapter in §3.3.1.5.

Irregular c' reflexes of PHEC *t'

The rest of the irregular reflexes of PHEC ***t'*** include a corresponding irregular **c'** as already seen in some of the EC parallels above. Therefore, the remaining paragraphs will elaborate on those correspondences where Burji has **c'**, Gedeo and Sidaama have **c'** or **t'**, and Hadiyya and Kambaata have **t'** (series 5–6).

The comparative sets [164] and [202] below provide the only consistent correspondences where both Gedeo and Sidaama exhibit initial **c'** in addition to Burji:

	[164]		[202]		[743]	
PHEC	***t'uf-**	'to close'	***t'all-**	'only, alone'	**-**	-
Burji	**c'ukk-**	'to close'	**c'alla**	'only, alone'	**-**	-
PNHEC	***t'uf-**	'to close'	***t'all-**	'only, alone'	***t'ork'**	'to chew'
Had	**t'uf-**	'to close'	**t'alla**	'only, alone'	**t'orok'-**	'be drilled'
Kam	**t'uf-**	'to close'	**t'alla**	'only, alone'	**t'ork'o**	'hole'
Sid	**c'uf-**	'to close'	**c'alla**	'only, alone'	**c'ork'a**	'to chew'
Ged	**c'uf-**	'to close'	**c'alla**	'only, alone'	**c'ork'a**	'to chew'
Cf. Oro	**c'ufa**	'close'	**c'alla**	'alone'	**-**	-

Other than [164], [202], and [743], the relevant sets show **t'** in Hadiyya and Kambaata, and usually also in Sidaama and Gedeo; only Burji consistently has **c'**. I therefore treat the Burji **c'** as a language-specific secondary reflex, not as evidence for reconstructing PHEC ***c'*** for the set as a whole, since the reflexes of PHEC ***c'*** are **c'** throughout HEC as discussed in the next section. This is evident in the following initial correspondence. This is evident in the following initial correspondences:

	[224]		[613]		[55]	
PHEC	***t'ur-**	'dirt'	***t'een-**	'rain'	***t'ut'-**	'black ant'
Burji	**c'ure**	'rubbish'	**c'eena**	'rain'	**c'uuc'c'ee**	'black ant'
PNHEC	***t'ur-**	'dirt'	***t'een-**	'rain'	***t'ut'-**	'black ant'
Had	**t'ura**	'dirt'	**t'eena**	'rain'	**-**	-
Kam	**t'uri**	'dirt'	**t'eena**	'rain'	**t'ut'i-**	'little ant'
Sid	**t'ure**	'dirt'	**t'eena**	'rain'	**t'ut'e, c'uunc'a**	'big black ants'
Ged	**t'ure**	'dirt'	**t'eena**	'rain'	**t'ut'e**	'small black ants'
Cf. Oro	**t'urii**	'dirt'	**-**	-	**t'iint'-**	'maggot'

The correspondence series with irregular **c'** is found in initial position only (cf. *Regular reflexes of PHEC *t'*).

Except [202][29] and [613], the other four examples above have parallels in the rest of the EC languages, with greater diversity in the reflexes, including the ones attested above and others, as presented below:

[29] PEOT has **kald'ay* 'alone, only' but the correspondences are also problematic when compared with PHEC [202] **t'alla* 'only, alone'.

[164]	t’	d’	c’, ʃ, j	Gloss
PHEC	***t’uf-**[30]			‘to close’
Oro			**c’ufa**	‘close’
Dir		**ɗuuʃ-**		‘cover, put a lid on’
Mos		**ɗuʃ-**		‘cover’
Kon		**ɗuf-**		‘close, enclose’
Bay			**c’uf-**	‘close’
Das			**ʃobónoʃ-**	‘cause it to close’
Som			**jufee**	‘insert a plug’

Two languages show a corresponding **t** and **d** for [164] Elmolo *tibʔe* ‘to bury’ and Yaaku *deps* ‘to close, to shut, to cover, to bury, to plug’. Sasse (1979, p. 20, 31) reconstructs PEC **d’₁uf-* ‘close, shut’ for the above correspondence.

[743]	t’	d’	c’, c, j	Gloss
PNHEC	***t’ork’**			‘to chew’
Burji		**d’okkomm-**		‘to chew (roasted grain or cane)’
Ts’am			**č’óx-**	‘milk’
Gaw			**c’oh-a**	‘to milk’
Yaaku			**c’aqau**	‘to chew’
Elm			**cak-**	‘to chew’
Som			**jaq**	‘suck something’

The above examples are problematic, including Burji [743] *d’okkomm-* ‘to chew’ and the EC parallels, which lack the PHEC ***r**.

[224]	t’	d’	ʃ, ɟ	Gloss
PHEC	***t’ur-**			‘dirt of body, clothing’
Oro	**t’urii**			‘dirt, filth, dirt of body, clothing’
Dir	**t’ur-á**			‘dust’
Arb	**t’uré**			‘dirty thing’
Afar		**ɗorók**		‘dirt’
Kon			**ʃur-eeta, ɟur-**	‘dirt, butter’

Different correspondences are found in Mositacha *ts’ura* ‘dust’ and Yaaku *durug* ‘dirt’. Again, Sasse reconstructs PEC **d’₁ur-* ‘dirt’ with a problematic ***d’₁** for such instances of **d’** corresponding to **c’**, **ʃ**, or **t’** (Sasse, 1982, p. 49; Sasse 1979, p. 29, 31). Such irregularity rather suggests this root may have undergone multiple phonetic shifts due to contact.

[30] According to Kießling & Mous (2003, p. 286), PWR **t’up* ‘smash’ is the cognate of PEC **d’₁uf-* ‘close, shut’, which is reconstructed here as PHEC [164] **t’uf-* ‘close’.

[55]	**t’**	**d’, ɖ**	**c’, j’**	**Gloss**
PHEC	***t’ut’-***			‘black ant’
Oro	**t’iint’-**			‘maggot’
Yaaku	**t’oroʔtɛ**			‘driver or safari ant’
Dah	**t’at’e**			‘small ant’
Dir		**ɗuuɗɗéʈ**		‘small black ants’
Mos		**ɗuɗe-ta**		‘ant’
Saho		**ɖuuɖe**		‘ant’
Afar		**ɗuuné**		‘ants’
Ren		**ɖúuɖe**		‘black ant(s)’
Maay		**ɖuuri**		‘ants’
Som		**ɖuuɖi**		‘a kind of louse’
Boni		**d’óói**	**j’úúrə, c’úúrə**	‘ant sp. (small)’

See Bayso [55] *tunce* ‘red ant’ and PWR **ts’ilaʔu* ‘small black ants (sp.)’.

Yaaku parallels in [224] and [55] still support a PHEC ***t’***. However, the evidence in the rest of SEC (Dullay) is not in consistent with PHEC ***t’***, this time as far as the parallels in [743] are concerned. The same sets show a consistent Somali correspondence **j**.[31] Excluding [55], **d’** is a less common correspondence compared to the previous sets.

Due to such irregularity, Sasse reconstructs the parallels of two of the above PHEC stems, i.e., [164] **t’uf-* ‘to close’ and [224] **t’ur-* ‘dirt of body, clothing’ with a problematic initial PEC ***d’**$_1$, from which Arvanites (1991, p. 136) extracts two PEC non-pulmonic consonants ***t’*** and ***c’***. For the present argument, however, the decisive issue is internal HEC reconstruction rather than PEC. At the PHEC level, these sets are better assigned to ***t’*** than to ***c’***, because their HEC correspondences are mixed. Additionally, for the regular PHEC ***c’*** sets discussed below, **c’** is retained across all five HEC languages.

The areal perspective might provide some answers for the potential sources of the irregular corresponding **c’** in HEC. Contact between these languages, Oromo, and Omotic languages is evident in certain cases. For instance, Burji, Gedeo, and Sidaama **c’** may be due to the heavy influence of Oromo, which explains why most of the cognates show a regular reflex reconstructable to PNHEC ***t’***, where the irregular **c’** is isolated and mainly in Burji.

[31] Somali **j** is the regular reflex of PEC ***c’*** (Arvanites, 1991, pp. 158–159).

	Oromo	→ Burji	→ Gedeo	→ Sidaama	Gloss
[164]	**c'ufa**	-	**c'uf-**	**c'uf-**	'close'
[202]	**c'alla**	**c'all-a**	**c'alla**	**c'alla**	'only, alone'
[25]	**c'ilee**, **c'ila-ttii**	**c'illee**	**c'ilaate**, **c'ila-tte**	**c'ilaatte**	'coal'
[13]	**hanc'ufa**	-	**hanc'ufa**	**-**	'saliva'

Beyond Cushitic, Bender (1988, p. 139) and Fleming (1988, p. 171) reconstruct Proto-Omotic ***ts'*** and Proto-South Omotic ***s'*** or ***ts'***, respectively. Thus, as an alternative to the above retention hypothesis, the sources of some Hadiyya/Kambaata **t'** and Burji **c'** might be Omotic **ts'** and **c'**, respectively, as in the following cases:

Omotic		→ Burji	→ Had	→ Kam	Gloss
Wolaytta **t'alaala**[32]	[202]	-	**t'alala**	**t'alla**	'only, alone'
Gofa **ts'alaala**		**c'all-a**			
Malo **c'uc'c'e**	[55]	**c'uuc'c'ee**	-		'louse', 'little ant'
Shinassha **ts'uts'a**				**t'ut'i-**[33]	'louse', 'black ant'
Wolayta **t'ugunta**	[267]	**t'unga**	**-**	**-**	'finger(nail)'
Bench **ts'ung**[34]					'claw'

The above examples include two of the four sets with irregular **c'**. Thus, if those two represent borrowings into HEC, the two reconstructions with PHEC ***t'*** (**t'all-* 'only, alone' and **t'ut'-* 'black ant') would no longer be valid. However, if they are inherited HEC words that were adapted due to contact, then the original initial consonant was ***t',*** as retained in Hadiyya and Kambaata.

The current dominant contact language is Amharic. Amharic loanwords with **s'**/**ts'** are transferred into HEC with **t'**. For instance, Treis (2008, pp. 24–25) confirms that in Kambaata "**/t'/** … regularly replaces the Amharic ejective fricative /s'/ in loanwords", e.g.,

[32] Lamberti & Sottile (1997, p. 529) propose the following changes: ***d'y-*** → ***t'-*** → ***ts'-*** for Gofa *ts'alaala* 'only'.
[33] In addition to Sidaama and Gedeo *t'ut'e* 'black ants'.
[34] Further parallels of [267] in Omotic include Koyra *ts'ugun-tse* (Sasse, 1982, p. 183), Shinassha *ts'ungu-ttsa*, Shekko *c'ungit'o*, and Malo *s'ungu-tsa* (Lamberti & Sottile, 1997, p. 528), all meaning 'fingernail'.

> Amharic *mäs'haf* 'book' → Kam *mat'aafá* (Treis, 2008, pp. 24–25), Burji *mat'aafi*, *mas'aafi* 'book' (Roba Dame & Wedekind, 2008, p. 60).
>
> Ethiosemitic **s'-ħ-f* 'write' (Lamberti & Sottile, 1997, p. 475) → Amharic *s'afä*, *t'afä* 'write' → Burji *t'aaf-*, Kam *t'aaf-*, Sid *t'aaf-*, Ged *t'aaf-* (Hudson, 1989, p. 171), Oro *t'aaf-*, *c'aaf-*, and *s'aaf-* 'write' (Lamberti & Sottile, 1997, p. 525).
>
> Amharic *s'omä* 'fast'→ Burji *t'oom-*.

Amharic *mäs'haf* 'book' is related to the Ethiosemitic stem **s'-ħ-f* 'write' (Lamberti & Sottile, 1997, p. 475). Oromo variants *t'aaf-*, *c'aaf-*, and *s'aaf-* 'write' are dialectal variations (Lamberti & Sottile, 1997, p. 525). The existence of variants with **s'** in Burji and Oromo indicates the recency of borrowing since **s'** is not a phoneme in either.

Amharic influence is ruled out for the sets under examination because of the lack of parallels, the lack of evidence to explain Amharic's contribution of the irregular **c'**, and the attested recency of contact with HEC. Therefore, if there are irregular **c'** and **t'** from multiple contact sources – Omotic languages (Lamberti & Sottile, 1997, pp. 253, 255, 258, 259, 260, 263) and Oromo – these scenarios must be proposed set by set after careful investigation.

Since Burji exhibits a tendency toward palatalization (§3.2.1), could we propose a palatalization rule ***t'** → **c'** to explain its irregular reflexes? There is no regular environment to justify all the occurrences of **c'** in Burji.

> [224] *c'ure* 'rubbish', [164] *c'ukk-* 'close with lid, to cover', and [55] *c'uuc'ee* 'black ant'
>
> [18] *c'eeji* 'blood' and [613] *c'eena* 'rain'
>
> [270] *c'ibba*, 'hundred'

In the above examples, Burji **c'** is found before **u** in [224] and [164], **e(e)** in [18] and [613], and **i** in [270]. Even if one counts **u** amongst the sounds that trigger palatalization, there are examples with a non-palatalized **t'** before **u** in Burji [267] *t'unga* 'finger, claw', despite the environment being similar to that of [224] *c'ure* 'rubbish' and [164] *c'ukk-* 'close with lid, to cover'. Additionally, there is a palatalized **c'** without a palatalization environment, e.g., [202] *c'alla*

'only, empty'. Thus, palatalization would not explain the irregular Burji correspondences. However, "harmony" can be proposed as an alternative hypothesis for some cases (and it is interesting that these are cases where a trigger exists for the previous palatalization hypothesis), such as in the reflexes of PHEC [18] **t'iig-* 'blood':

> Burji *c'eeji*, Gaw *c'eeq-t-e*, Ts'am *c'egde*, Boni Kilii *c'íìk* 'blood', Kam *k'egu*, *k'egi*, Gollango *k'eq-te* 'blood' (Bender, 2020, p. 2).

A simpler explanation of the above examples is that they borrowed from each other, but they remain irregular within the broader EC context.[35] Therefore, via obstruent harmony,[36] Kambaata and Gollango might have assimilated C_1 **t'** to the velar in C_2, while Burji might have assimilated C_1 **t'** to the palatalized **g** → **j** in C_2 (§3.2.1). Then, the initial **k'** triggered the glottalization harmony of ***g** to **q** in Gawwada *c'eeq-t-e* and Gollango *k'eq-te* 'blood'. This obstruent harmony scenario is more plausible than the other explanations proposed.

The contact-induced **c'** reflexes of PHEC ***t'** discussed above differ from the glottalized **c'** arising through fusion in PNHEC (§3.3.1.4). The fusion process creates glottalized (secondary non-pulmonic) consonants from obstruent-glottal sequences rather than from inherited proto-phonemes or from donor languages. This distinction matters because fusion-derived **c'** is regular and applies across the NHEC languages wherever the phonological conditions are met, whereas contact-induced **c'** appears sporadically in specific lexemes.

PHEC *c'

After establishing the contact scenarios that explain irregular **c'** reflexes of PHEC ***t'** in certain HEC forms, the discussion now turns to examining lexemes that show regular **c'** correspondences across all five HEC languages.

---

[35] As expected, most of the (Lowland) EC correspondences show **t'**/**d'**: [18] Had *t'iiga* 'blood', Oro *d'iiga* 'blood', Dir *d̃iik* 'blood, sap', Mos *d'iika* 'blood', Kon *d̃iika* 'blood', Ren *ɖíig* 'blood', Som *ɖiig* 'blood', Maay ɖiig 'blood', Boni *d'íìk* 'blood'; Arb *d'iik'* 'blood', Dah *d̃iiga* 'blood', cf. PWR **ts'eedee* 'blood'.

[36] A broader hypothesis of "obstruent harmony" (between C_1 and C_2) might alternatively explain other instances of irregular correspondences that have been given various other interpretations in this study. For instance, the problematic glottalization of the labial: PNHEC [49] **k'uub-* 'egg' in Kam *k'up'p'a-ta* 'eggs', Sid *k'uup'p'e* 'eggs, egg', Ged *k'uup'e*; *k'uup'p'e* 'eggs, egg', and the full assimilation in Had *k'uunk'a* 'egg'. Obstruent harmony accommodates sibilant, velarization, glottalization, voicing, and other forms of assimilation (§2.5.2).

These regular correspondences support the reasonable reconstruction of PHEC ***c'*** as a distinct proto-phoneme.

PHEC ***c'*** or ***tʃ'*** is a non-pulmonic post-alveolar affricate. The reconstruction of PEC ***c'*** Ehret (1987) and Arvanites (1991, pp. 158–159) is supported by evidence across HEC languages and correspondences from East Cushitic, such as Sidaama **c'**, Oromo **c'**, Somali **j**, and Afar **d'**. Ehret (1987, p. 34) reconstructs **c'** for both PEC and PC, with consistent reflexes in Yaaku and Oromo **c'**. This section analyzes the regular and irregular reflexes of PHEC ***c'***, the EC correspondences, and potential external influences.

The reconstruction of PHEC ***c'*** is based on regular correspondences in initial and non-initial positions and cognates found across the HEC languages. The examples below present the primary evidence that supports the reconstruction of PHEC ***c'*** in initial position:

	[516]		[186]		[348]	
PHEC	***c'il-***	'to defecate'	***c'ub-***	'dip in, immerse'	***c'ii?-***	'bird'
Burji	**c'il-**	'to defecate'	**c'uub-**	'dip in, immerse'	**c'iid'aa**	'bird'
PNHEC	***c'il-***	'to defecate'	***c'ub-***	'dip in, immerse'	***c'ii?-i-ʃʃ-***	'bird'
Had	**c'ir-**	'to defecate'	**c'ufc'ufa a?-**	'dip in, immerse'	**c'ii?-cco**	'bird'
Kam	**c'in-**	'to defecate'	**c'uf a?**	'dip in, immerse'	**c'ii-ccu-**	'bird'
Sid	**c'il-**	'to defecate'	**c'uu?i**	'dip in, immerse'	**c'e?-icco, c'i?a**	'bird'
Ged	**c'il-**	'to defecate'	**c'uup'-**	'dip (in)'		'bird'
cf. Oro	**-**	-	**c'uup'-**	'dip, baptize'	**-**	-

In [186], the change of ***b*** to **f** in Hadiyya and Kambaata and **p'** in Gedeo is explained via morpheme boundary alternations in §3.3.1.1. There is a subsequent reduplication of the same root in Hadiyya. The **uu** vowel doubling in Burji and Sidaama is likely due to contact with Oromo or internal alternation processes. In set [516], the **n** in Kambaata [516] *c'in-* 'to defecate' is irregular.

For additional examples, see PHEC [516] **c'il-* 'to defecate' with [517] **c'iil-* 'excrement', [186] **c'ub-* 'dip in, immerse', with [286] **c'ob-ʔ-* 'to drip, to leak' and [447] **c'ubb-* 'sin', [348] **c'ii?-* 'bird' with [226] **c'iwʔ-*, **c'awʔ-* 'chick(en)'.

There are three more examples of initial ***c'** reconstructable to PNHEC: [21] **c'ak'as-* 'to listen, spy', [416] **c'in-*[37] 'to squeeze', and [44] **c'aʔ* 'to clean'.

In non-initial position, additional evidence for PHEC ***c'** is found in stems that retain regular reflexes as follows:

	[580]		[491]		[807]	
PHEC	***boc'-**	'to carve, split'	***fac'-**	'to cut, break'	**-**	-
Burji	**boc'-**	'carve'	**fac'-**	'to cut'	**-**	-
PNHEC	***boc'-**	'to carve, split'	**-**	-	***inc'-**	'to chew, twist'
Had	**boc'-**	'to carve, split'	**fac'ite**	'burst of swelling'	**inc'iic'-**	'twist'
Kam	**boc'(c')-**	'snap (to break)'	**-**	-	**inc'iic' c'-**	'twist'
Sid	**boc'-**	'to carve, split'	**fac'i**	'break a branch'	**hiinc'i**	'chew'
Ged	**boc'-**	'cut, chop'	**-**	-	**inc'a**	'chew'
cf. Oro	**-**	-	**-**	-	**(h)anc'-**	'chewing gum'

All three of the above sets support the reconstruction of ***c'** in non-initial consonant in PHEC, with reflexes in Burji and PNHEC exhibiting regular correspondences. Set [580] has the Burji parallel *annij-* 'to chew', whose status as a cognate is doubtful because of deviant semantics, but it could be argued to be an older form with **c'** in PNHEC resulting from **ʃ** and **ʔ** fusion as discussed in §3.3.1.4. Other examples of non-initial HEC ***c'** include PHEC [377] **c'anc'-* 'to make noise' and PHEC [162] **fic'c'-* 'to scatter'. Interestingly, most of the above non-initial correspondences exhibit the sequence **nc'**, including an epenthetic **-n-** (§3.4.3).

Many examples of PHEC ***c'** align with EC parallels, providing wider supporting evidence for PHEC ***c'**:

> PHEC [186] **c'ub-* 'dip in, immerse': Oro *c'uup'a* 'dip, baptize', *c'up'-*, *c'ub-* 'dip in', Dir *c'ap-* 'be wet', Mos *c'abiso*, *c'ubolayca* '(be) wet, be corrupt', Kon *ʃopɕ* 'to dip into water', Arb *c'um-* 'wade', Ts'am *c'ab'-* 'to rot, to become wet'.

[37] See Som *ɖiiji* 'milk a little, squeeze a trickle from', *diiq* 'trickle (of milk)'.

PHEC [226] **c'iwʔ-*, **c'awʔ-* 'chick(en)': → Oro *c'uc'-oo*, *c'uc'-ii*, *wic'-oo*, *wic'ii* 'chick', Dir *c'aac'ut-ét*, *c'uuc'-éṭ* 'chick, (small) chicken', Kon *ʃaaʃʃ-eeta* 'bird sp.', Bay *c'uc'ute* 'chicken', Boni *c'eekúù* 'bird sp.', Gaw *č'uč'uw-e* 'chicks'.

PNHEC [416] **c'in-* 'to squeeze': Oro *c'unfa* 'squeeze'.

PNHEC [44] **c'aʔ* 'to clean': Dir *c'ik'* 'wash (tr), clean from a curse of witchdoctor', Mos *c'ik'* 'wash', Kon *ʃaɕ-* 'wash'.

PHEC [84] **c'ab-* 'to cut, sharpen': Oro *c'ab-a* 'break, be broken', Ren *jebic*, *jébi* 'break (something)', Som *jebi* 'break', Maay *j'ibiyow*, *j'abow*, *j'ab-* 'break', Boni *c'ibiiy-*, *j'ibiiy-*/*c'iiy-* 'to break' (cf. PSC **tʸ'uf-* 'to trickle' and PC **c'abk'-* 'to break up, break apart' (tr) (Ehret, 1987, p. 40)).

PNHEC [21] **c'ak'as-* 'to listen, spy': Arb *tʃ'ak'as-* 'listen' 'hear'.

PHEC [377] **c'anc'-* 'to make noise': Oro *c'ac'c'isa* 'fight back, snap back', Dir *c'aac'c'-* 'discuss (noisily)', Mos *c'ac'* 'quarrel', Kon *ʃaaʃʃ-* 'make a high pitched cry', Som *jaar-i* 'difference, contention (in a contest)'.

PHEC [580] **boc'-* 'to carve, split (wood)': Arb *botʃ'-* 'beat, hew', Bay *boc'i* 'beat, hew', Yaaku *pilc'* 'small sticks of firewood' (cf. Bilin *bɐc'bɐc'-*, *bɐt'bɐt'-* 'shake' (Lamberti & Sottile, 1997, p. 314), Som [321] *buji* 'pierce', Maay *buj'iyow* 'pierce', Boni *bac'-iy-*, *baj'-iy* 'to pierce').

PHEC [491] **fac'-* 'to cut, break': Oro *fac'c'ee* 'fighting with a stick or sword'.

PNHEC [807] **inc'-* 'to chew, twist': Oro *(h)anc'-aa* 'chewing gum', Som *ɦanj-o* 'chewing gum', Ren *ɦánja* 'gum from certain trees', Mos *ʔɑnc'uhiso* 'chew'.

The PNHEC stem **c'aac'urr-*, **c'iic'oorr-* 'chick(en)' with **rr** in NHEC formally aligns with the following EC parallels Oro *c'irrii* 'bird', Gaw *c'irraʔ-itt-e* 'red-billed oxpecker', Ts'am *c'uruq'e* 'bird sp.', and Arb *tʃ'aaraté* 'oxpecker'. However, those parallels semantically align with PHEC [348] **c'iiʔ-* 'bird'. Arbore probably has two archaic stems *tʃ'ak'as-* 'listen' and *d'eg(a)y-* 'hear', but one parallel in PNHEC [21] **c'ak'as-* 'to listen, spy'. According to Arvanites (1991, p. 172), in Somali [377] *jaar-i* 'difference, contention (in a contest)', **r** is

from ***d'***. Another root discussed above, PHEC [516] **c'il-* 'to defecate', has no EC or areal parallel.

The number of sets with regular correspondences in both initial and non-initial positions in HEC, combined with the numerous EC parallels, supports the reconstruction of PHEC ***c'***.

Three potential contact sources could explain irregular **c'** in HEC: Oromo borrowing, Omotic substrate influence, and Amharic loanwords. Each requires a separate evaluation.

Despite their regular correspondence of ***c'*** within PHEC, some of the above forms with Oromo parallels might nevertheless represent loans into HEC languages based on form and semantics, such as [186] or irregularity of correspondences with the rest of HEC, as in [162].

	Oromo	→ Burji	→ Ged	→ Sid
[186]	**c'uup'-** 'dip, baptize'	-	**c'uup'-** 'dip (in)'	-
[377]	**c'ac'c'isa** 'fight back, snap back'	**c'anc'-** 'to rant'	**c'anc'-** 'shout, cry'	**c'anc'-** 'noise, shouting'
[169]	**k'unc'ee** 'bark'	**k'unc'e** 'bark'	**k'unc'e** 'bark'	**-**
[162]	**fac'c'aʔa** 'be scattered, sown', **fac'c'aasa** 'scatter, sow'	**fac'c'-** 'scatter'	**fac'aasa** 'scatter'	**-**

The forms with cognates within the rest of HEC and EC might represent cases where the existing cognate has been replaced or include double reflexes as a result of the loans. See [169], [367], and [476] for more examples.

Contrary to the retained ***c'*** proposed by Arvanites (1991) and Ehret (1987), Sasse (1979, p. 30) argues that **c'** in the HEC languages predominantly appears in loanwords, with two or three autochthonous instances that lack cognates outside HEC. Sasse's statement overgeneralizes and only aligns with instances where **c'** is irregular across HEC languages, which may reflect contact influences. The remaining paragraphs of this section review such irregular correspondences in the above examples with their EC parallels. Notable cases include:

PHEC [186] **c'ub-* 'dip in, immerse': Afar *d'uum-* 'sink', Som *ku ɖumbi* 'plunge something into water, immerse', Gaw *k'ap-* 'to rot (of dead animals)' (cf. Dah *t'up'-* 'to leak').

PHEC [226] **c'iwʔ-*, **c'awʔ-* and PNHEC **c'aac'urr-*, **c'iic'oorr-* 'chick(en)': Mos *cu-cute* 'chick', Som *ɖawɖawle* 'bird sp., woodpecker'.

PNHEC [416] **c'in-* 'to squeeze': Som *ɖiiji* 'milk a little, squeeze a trickle from' and *diiq* 'trickle (of milk)', Arb *d'angalay-* 'gush, pour out'.

PNHEC [44] **c'aʔ* 'to clean': Ren *ɖixa* 'wash (something)', Som *ɖaq-* 'wash', Maay *ɖiɣow*, *ɖiq* 'to wash', Boni *d'ak'aʔ*, *d'aʔ-* 'to wash', Arb *d'ik'-* 'wash', Das *ʃik*, *gĩk-* 'to wash', Elm *d'ik* 'to wash', Yaaku *t'ok'-* 'wash'.

PHEC [84] **c'ab-* 'to cut, sharpen': Dir *k'ep-* 'break', Mos *k'eɓ-* 'break', Kon *k'ep-* 'break', Bay *ebis-* 'break' (tr), Arb *k'eb-* 'break', Das *gɛ́b* 'to break'.

PNHEC [21] **c'ak'as-* 'to listen, spy': Oro *d'agaya* 'hear', Dir *ɗakay-* 'hear', Mos *ɗakay*, *ɗak'a* 'listen, hear', Kon *ɗak-ay-* 'hear', Saho *-eɖeg-* 'know, recognize', Afar *ɖag* 'auricular region', Ren *ɖaaga* 'hear', Som *ɖegeyso* 'hear, listen to', Maay *ɖeeɣow* 'hear, listen', Boni *diŋésio* 'listen', Elm *d'ekait*, *d'ey* 'to hear', Yaaku *dek-* 'hear'.

PHEC [580] **boc'-* 'to carve, split (wood)': Oro *bat'bat'* 'shake', Bilin *bɐc'bɐc'-*, *bɐt'bɐt'-* 'shake' (Lamberti & Sottile, 1997, p. 314).

PHEC [491] **fac'-* 'to cut, break': Mos *faats'o* 'branch (of tree)'.

For PHEC [186] **c'ub-* 'dip in, immerse', Sasse (1979, p. 30) reconstructs PEC **d'₁ub-* 'dip in'. Parallels of PHEC [226] **c'iwʔ-*, **c'awʔ-* and PNHEC **c'aac'urr-*, **c'iic'oorr-* 'chick(en)' are attested in Saho as a loanword *s'us'uut* 'chicken' from Tigrinya/Arabic, as claimed by Esayas Tajebe (2015, p. 30), with no matches in Omotic. It is also likely an onomatopoetic root. In [21], Arbore shows both **d'** and **c'** variants, potentially reflecting a distinction between two separate roots: one with **c'** (← ***c'**) and one with **d**' (← ***t'**). If other EC languages have merged these roots, the two reflexes here would represent regular developments from their respective proto-forms. Sasse (1979, p. 17) reconstructs PEC **d'eg-/*dog-* 'to hear' (cf. PC **dleg-* 'to listen' (Ehret, 1987, p. 83)). In [580] **boc'-* 'to carve, split (wood), Oro *bat'bat'* 'shake', has been transferred into Gedeo as *bat'a* 'split open, spread'. This root has parallels in Bilin *bɐc'bɐc'-* and *bɐt'bɐt'-* 'shake' (Lamberti & Sottile, 1997, p.

314). The existence of both **c'** and **t'** in the parallels complicates the cognacy relationship.

The most common correspondences in the above examples are **d'** with a few cases of **k'/ɠ**, **d**, and a single attestation of **c**, **ʃ**, **ø**, **ts'**. Establishing a regular palatalization rule to explain **t'/d'** → **c'** is difficult,[38] but as Arvanites (1991, p. 130) proposes, it might be an areal change.[39]

Several sets show regular **c'** correspondences across all HEC languages, but multiple factors suggest they are Amharic loanwords: (1) limited or absent EC cognates outside HEC, (2) cultural domains (clothing, weapons, stimulants), and (3) close phonological matches with Amharic forms.

Amharic		**P(N)HEC**		**→ HEC**
c'ork'a 'rag'	[220]	**c'ark'-*	'rags, cloth'	Burji *c'ark'aa*, Ged *c'ark'e*, Had, Kam *c'ark'a*
c'ube 'dagger'	[544]	**c'ub-*	'dagger, knife'	Had *c'ubeʔe*, Kam *c'up'i-ta*, Ged *c'uup'e*, Sid *c'uube*
c'amma 'shoe'	[81]	**c'amm-*	'shoes'[40]	Ged *c'aamma*, Sid *c'amma*, Had *c'aammaʔa*, Kam *c'amma*
c'at 'chat (Catha edulis)'	[513]	**c'aat-*	'chat (Catha edulis)'	Had, Kam *c'aata*, Sid, Ged *c'aate*
c'äbt' 'gonorrhea'	[490]	**c'obt'-*	'gonorrhea'	Burji *c'ooddoo*, Had *c'abt'oʔo*, Kam *c'obt'a*

Despite the complicating factor of Amharic loanwords showing regular **c'** correspondence, the overall pattern for PHEC (i.e., inherited) ***c'** is that it has regular correspondences in HEC, it has more (and mostly regular) parallels in (East) Cushitic, and the reconstructed words have fewer parallels in the neighboring non-Cushitic languages. The attested Oromo loans appear to involve double reflexes, i.e., reflexes of original PHEC forms alongside borrowed reflexes in the cognate HEC and Oromo lexemes.

---

[38] Is it possible that, in those cases, ***c'** irregularly changed to **k'**?

[39] According to Arvanites (1991, p. 130, 157), "[n]ot to be overlooked is the fact that *t' sometimes is perplexing!; Correlatable with *k'. The way that this is accomplished via regular phonetic change involves palatalization (a common areal change involving alveolars) of *t' to c'."

[40] Reconstruction [81] actually is **kob-ʔ-* 'shoe, sandal' but **c'amm-* is proposed as an additional provisional reconstruction. Amharic *c'amma* 'shoe' is from Lamberti & Sottile (1997, p. 331).

The reconstruction of PHEC ***c'** thus rests on regular correspondences across all five HEC languages, supported by wider EC parallels. This non-pulmonic ***c'** must be distinguished from glottalized ***c'** arising through the fusion of sibilants with the glottal stop in NHEC. The fusion process (§3.3.1.4) creates **c'c'** geminates wherever **ʃ** or **s** precedes **ʔ**, yielding forms that superficially resemble inherited ***c'** reflexes but follow a different developmental path.

Summary

The reasonable reconstruction of PHEC ***t'** and ***c'** as distinct proto-phonemes rests on systematic correspondence patterns across all five HEC languages, supported by extensive attestation in both word-initial and non-initial positions. Each phoneme shows regular reflexes: ***t'** yields **t'**, while ***c'** consistently yields **c'** across HEC. The complexity arises not from the robust regular correspondences themselves but from explaining why subsets of lexemes show irregular reflexes.

The irregular **c'** and **d'** reflexes in the ***t'** correspondence set reflect multiple contact scenarios rather than undermining the reconstruction. First, direct borrowing from Oromo (which has **d'**) introduces **d'** into specific HEC lexemes, particularly in Burji, Gedeo, and Sidaama, due to sustained contact. Second, analogy and harmony processes operating within HEC can produce secondary **c'** from inherited ***t'** in specific phonological environments. Third, earlier contact with Omotic languages contributed additional irregular patterns with **c'** or **t'**. Crucially, these contact-induced irregularities are identifiable precisely because the regular inherited patterns are so clear.

The reconstruction of ***c'** as distinct from ***t'** receives further support from three independent lines of evidence: (1) numerous EC cognates showing systematic ***c'** correspondences; (2) the semantic distribution, wherein ***c'** reconstructions cluster in basic vocabulary unlikely to be wholly borrowed; (3) the phonological naturalness of the contrast, which exists throughout Cushitic and reflects the articulatory distinction between apical (**t'**) and post-alveolar (**c'**) non-pulmonics.

This analysis demonstrates the methodological approach outlined in §1.5: treating contact and inheritance as a continuum rather than binary alternatives. Some lexemes inherit ***t'** regularly, others inherit ***c'** regularly, and still others show complex histories involving borrowing, analogy, or secondary processes. This nuanced approach provides a more historically realistic reconstruction than forcing all data into rigid categories.

2.1.3 Labials

Highland East Cushitic has four labial consonants: the stop ***b**, the nasal ***m**, the fricative ***f**, and the glide ***w**. This section examines the ***b**, ***m**, and ***f**, while the labial glide ***w** is discussed with the other glide ***y** in Section 2.1.4. Where ***b** lenition involves an intermediate ***w** stage, only the necessary relevant evidence is presented here; the full discussion of ***w** reflexes follows in the PHEC ***w** section. Their historical development reflects three key processes that shaped their modern reflexes across the daughter languages:

a. Spirantization of intervocalic ***b** to [**β**] in Burji, Sidaama, and Gedeo
b. Lenition of intervocalic ***b** to ***w**, sometimes with subsequent loss
c. Loss of intervocalic ***w**

These three processes form a feeding relationship, highlighting the historical tendency to weaken labials (***b** and ***w**) in intervocalic position. While the fricativization is allophonic, the lenition of ***b** appears to be often triggered by more recent boundary alternation processes compared to that of ***w**, though they may have operated simultaneously at an earlier stage.

PHEC *b

PHEC ***b** is securely reconstructed based on evidence in both initial and non-initial positions.

	[34]		[9]		[183]	
PHEC	***baʔ-**	‘be lost’	***arrab-**	‘tongue’	***gib-**	‘to refuse, hate’
Burji	**be(y)-, bay-**	‘be lost’	**arraaba**	‘tongue’	**jibb-**	‘to hate’
PNHEC	***baʔ-**	‘be lost’	***arrab-**	‘tongue’	***gib-**	‘to refuse’
Had	**biʔ-**	‘to be lost’	**allabo**	‘tongue ’	**gib-**	‘flee’
Kam	**baʔ-**	‘be lost’	**arrabi-ta**	‘tongue’	**gib-**	‘refuse’
Sid	**baʔi**	‘perish’	**ʔarrabo, arrawo**	‘tongue’	**gib-**	‘refuse’
Ged	**baʔ-**	‘be lost’	**arrabo**	‘tongue ’	**gib-**	‘refuse’

In [183], the geminate **bb** in Burji might be due to the influence of Oromo *jibba* ‘dislike’.[41]

[41] “The Burji and Gedeo words could be purely borrowings from Oromo [jibba], but perhaps the Oromo word, presumably a distant cognate, has just influenced these forms” (Hudson, 1989, p. 7). Additionally, in Sidaama, “assimilation takes place when

Additional examples include PHEC [444] **bagaz-* ‘spear’, PNHEC [460] **beet-* ‘son, child’, PHEC [273] **biin-* ‘mosquito, fly (sp.)’ (see Amh *bimbi* ‘mosquito’), PHEC [407] **ebel-*, **ebal-* ‘so-and-so’ (likely compound word), PNHEC [112] **bub-* (intr) ‘burn’.

Burji and Gedeo show spirantized variants **b ~ β** intervocalically (Sasse, 1982, p. 16; Eyob Kelemework, 2015, p. 34), while Sidaama exhibits both **β** and **w** variants of **b** intervocalically (Yri & Pepper, 2019; Anbessa Teferra, 2000, p. 26; Hudson, 1989, p. 12; Bender, 2020, pp. 116–117). This variation is reflected in different transcription practices across sources: Yri & Pepper (2019) use phonemic /**b**/ and orthographic <**w**>, Gasparini (1983) and Indrias Xa’miso et al. (2007) have [**w**].[42] The data is cited in this study without modification in this regard. The spirantization ***b** → [**β**] in these three languages represents an initial stage of lenition without complete loss when stem-final **b** is followed by a vowel through morphophonological processes. Examples from Sidaama include:

b ~ β: *ʔaaba ~ aaβa* ‘father’
b ~ w: *ʔarrabo ~ arrawo* ‘tongue’
duub- ‘be satisfied’, *duuwí* (perf) ‘He was satisfied’[43]

The nouns are followed by the terminal vowels **-a** and **-o**, and the verb is followed by the perfective **-í**. In [252] *hasaaw-i ~ hasaab-i* ‘speak’, even an original **w** has an alternant **b**.

A distinct lenition process affecting intervocalic ***b** predates the synchronic spirantization described above. This older rule resulted in the complete loss of ***b** in intervocalic environments:

ii. Lenition of intervocalic ***b**
***b → *w /V_V**

---

the suffix /-tú/, 3rd f. / 3rd pl. perfect marker, is attached to stem-final obstruents” such as in *gib-* ‘refuse’ *gibbú* (perf) ‘she/they refused’ (Anbessa Teferra, 2000, p. 21).

[42] “It is rather unfortunate that the decision makers chose to write the allophone [β] of /b/ as <w> intervocalically rather than <b>, the more so as the intuitively correct <b> is maintained word initially and <bb> for the geminated one word medially. The phoneme /w/ occurs word initially and in geminated form it contrasts with /bb/ word medially as /ww/. /w/ has lip rounding in all positions while [β] (allophone of /b/) is only rounded next to vowels with lip rounding, viz. /o/ and /u/. And this environmentally induced lip rounding is probably the cause of the unfortunate choice” (Yri & Pepper, 2019).

[43] Data from Anbessa Teferra (2000, p. 26) and Hudson (1989, p. 7).

Evidence for this historical process can be found in Burji, which retains the ***b** in consonant clusters, such as **bd**, while NHEC languages show a complete loss of this sound in intervocalic positions. The lost ***b** can be reconstructed based on three lines of evidence: Burji cognates, East Cushitic cognates, and NHEC vowel correspondences.

> PHEC [115] **k'ab y-t-* 'become cold' → Burji *kabb-* 'cold, cool, become cold'; PNHEC **k'iid-* 'become cold' → Had, Sid, Ged *k'iid-* 'cool, become cold', Kam *gid-* (cf. Oro *k'abb-an-* 'become cool (liquid)').

Beyond the assimilation in Burji of ***bd** to **bb**, the corresponding vowel patterns and the final consonant in NHEC support the reconstruction of the verbalizing agentive denominative ***y-t** (§4.2.5.3). The ***b** in this root is additionally attested in other EC cognates (PEC **k'ab-*), with the final **d** likely deriving from the fossilized ***-t** in HEC. The loss of ***b** in NHEC can be reconstructed in three phases:

(a) Creation of an intervocalic environment via the verbalizer vowel **i** (§4.2.5.2): **k'ab-i-d-*
(b) Lenition of ***b** to **w**: **k'awi-d-*
(c) Loss of intervocalic ***w** and merger of ***a** and ***i** into long **ii**, yielding PNHEC **k'iid-*

While this reconstruction accounts for most of the developments, the initial velars in Burji (**k**) and Kambaata (**g**) remain problematic in light of their correspondence with non-pulmonic **k'** in the rest of HEC.

Hudson (1989, p. 12) notes that there is no evidence for reconstructable intervocalic ***b** within roots in HEC. Synchronically, intervocalic **b** appears only in borrowed words and a few retained exceptions. This is exemplified by [349]**gabal-* 'border, limit', which shows both irregular intervocalic **b** in all HEC and irregular **l** in Hadiyya (where **r** would be expected §2.1.1.3). The exceptions include PHEC [214] **t'aban ʔ*, **t'aban ʔ-s-* 'to slap' (with loss in Sidaama *d'aʔa assa* 'to slap (the face)' only via lenition ***ø** ← ***w**) and PHEC [606] **gubeed-* 'thigh' (see Oro *gudeeda*). These represent arguable cases of exceptional retention, where morphological boundaries or lexical factors may have blocked the otherwise regular intervocalic lenition process. The limited number of such exceptions (compared to the overwhelming pattern of loss) strengthens rather than undermines the reconstruction of the regular rule.

Non-initial ***b** is not affected by the lenition rule when it occurs in a geminate or in a cluster, as demonstrated in the following cognate sets:

	[270]		[3]		[95]	
PHEC	***t'ibb-**	'hundred'	***aabb-**	'father'	**ʃomb-**	'lung'
Burji	**c'ibba**	'hundred'	**aabboo**	'father'	**ʃombi**	'lung'
PNHEC	***t'ibb-**	'hundred'	***aabb-**	'father'	**-**	-
Had	**t'ibbe**	'hundred'	**aabba**	'father'	**-**	-
Kam	**t'ibbi-ta**	'hundred'	**aabba**	'father'	**-**	-
Sid	**t'ibbe**	'hundred'	**ʔaaba**	'father'	**ʃombo**	'lung'
Ged	**(d'ibba)**	'hundred'	**abbo**	'uncle'	**ʃombo**	'lung'

More complex correspondences emerge in examples with vowel alternations and ***lb** clusters.

> PHEC [4] **gulib-*, **gilub-* 'knee' → Burji *gilba* 'knee'; PNHEC **gulub-* → Had *gurubbo*, Kam *gulubi-ta*, Sid *gulucco*, Ged *gulubo* (cf. Oro *jilba* 'knee', PEC **gilb* → **gilib* → **gilub*).

The correspondence patterns and vowel reflexes require careful analysis. While rounding of V_2 (***i** to **u** before **b**) followed by V_1 assimilation might explain the NHEC forms, the comparative evidence points to a different reconstruction. Three independent lines of evidence converge on **gulib-* as the PHEC form: (1) the lack of palatalization in the Burji form indicates V_1 was already ***u** at the proto-level, (2) the Oromo cognate *jilba* preserves the expected ***i** in V_1 position before ***l** (consistent with PEC **gilb-*), and (3) the NHEC vowel assimilation (**gulib-* → **gulub-*) follows attested patterns of labial-conditioned rounding. This reconstruction demonstrates the application of the comparative method, where multiple phonological processes operating at different chronological stages must be disentangled to arrive at the most parsimonious proto-form.

Some HEC languages exhibit problematic correspondences involving the sounds **b**, **m**, and **mb**. The nasal variants **m**/**mb** appear specifically in Burji, Gedeo, and Sidaama examples:

[514] Burji *amb-* ‘bring’: PHEC **ebb-* ‘to bring’ → PNHEC **ebb-* ‘to bring’ → Had and Kam *eeb-* ‘bring, marry (a woman)’, Sid *ʔabb-* ‘bring’, Ged *eʔiʃa* (← **ew-iʃ-*)[44] ‘bring in’.

[495] Gedeo *lumo* ‘big’: PHEC **lab-* ‘many, big, much’ → Burji *laboo*, Had *lob*, Sid *lobo*.

[550] Sidaama *ʔibiicco* (f) (Yri & Pepper, 2019); *imiiwe* (sg), *iwiiwe* (pl) (Indrias Xa’miso et al., 2007, p. 273); *ibi-cco*, *ibiibe* (Hudson, 1989, p. 94) ‘louse, lice’: PHEC **ibb-* ‘louse’ → Burji *ibba*; PNHEC **ib-*, **ib~ib-* ‘louse’ → Had *ibiiba*, Kam *ibiibi-ta*; *ibiibe* ‘louse’, Ged *ibbe*.

These irregular correspondences can be explained through epenthetic **-n-** (§3.4.2), as evidenced by cognates in other HEC and EC languages (e.g., [514] Jiddu *eeb-* ‘bing’ and Das *ʔeeb-* ‘bring’). The following development pathway can be proposed:

[514] Burji *amb-*	‘bring’	← **a-n-b-*
[495] Gedeo *lumo*	‘big’	← **lu-n-b-*
[550] Sidaama *imiiwe*	‘lice’	← **i-n-biib-*

Based on this interpretation, **m** represents an inserted and assimilated **-n-** before the labial, found as the cluster **mb** in [514] but reduced to **m** in [495] and [550].

In summary, PHEC ***b** shows the following key patterns: preservation in initial position, occurrence in non-initial position primarily as geminates (some potentially resulting from old assimilation of ***b** and another consonant (§3.1.1)) or in clusters, loss of intervocalic single ***b** via the glide ***w** following a pattern common to other EC languages, and glottalization and occasional fricativization in final position when followed by voiceless consonants (**ʔ**, **s**, **ʃ**, or **c**) in morpheme boundary alternation processes (§3.3.1).

PHEC *m

The labial nasal ***m** is securely reconstructed with evidence from regular reflexes across the HEC languages. Unlike the alveolar nasal ***n** (see

[44] Comparable with the ***b** loss in Sid *d’aʔa assa* ‘to slap (the face)’ ← PNHEC [214] **t’aban ʔ-s-* ‘to slap’.

§2.3.1.4), ***m** occurs regularly in both initial and non-initial positions, as demonstrated in the following cognate sets:

	[54]		[72]		[48]	
PHEC	***min-**	‘house’	***lam-**	‘two’	***tum-**	‘to pound’
Burji	**mina**	‘house’	**lám-a**	‘two’	**tum-**	‘to hit/churn’
PNHEC	***min-**	‘house’	***lam-**	‘two’	***tum-**	‘to pound’
Had	**mine**	‘house’	**lamo**	‘two’	**tum-**	‘to hit/strike’
Kam	**mini**	‘house’	**lamu**	‘two’	**tum-**	‘to hit/strike’
Sid	**mine**	‘house’	**lame**	‘two’	**tum-**	‘to hit/strike’
Ged	**mine**	‘house’	**lame**	‘two’	**tum-**	‘to pound’

Further examples include PHEC [68] **malaɦ-* ‘pus’, [268] **mur-* ‘to cut’, [547] **am-* ‘mother’, [675] **k’oom-* ‘calabash’, and [391] **kum-* ‘thousand’.

While ***m** shows general stability, it undergoes important morpheme boundary alternation processes in specific environments. Most notably, stem-final ***m** exhibits partial assimilation to **n** before alveolar-initial suffixes (see §3.1.1.4 for a detailed discussion). This process remains productive in modern HEC languages, as evidenced by Sidaama’s synchronic assimilation of stem-final **m** to **n** before **t** (Kawachi, 2007, pp. 49–50).

PHEC ***m** also plays a crucial role in the morphological system in the form of the reconstructable passive suffix **-am** (see §4.2.1). The presence of the labial nasal **m** in this morpheme is consistent across HEC, along with its regular phonological alternations, which supports its reconstruction for PHEC.

PHEC *f

The reconstruction of the labial fricative ***f** for PHEC is secure and supported by regular initial and non-initial reflexes across all HEC languages. The following cognate sets demonstrate these stable correspondences:

	[167]		[70]		[40]	
PHEC	***fayy-**	‘healthy’	***afoʔ-**	‘mouth’	***tuf-**	‘to spit’
Burji	**fayya-n-ga**	‘healthy’	**afeyi**	‘mouth’	**tuf-**	‘to spit’
PNHEC	***fayy-**	‘healthy’	***afoʔ-**	‘mouth’	***tuf-**	‘to spit’
Had	**fayyaʔa**	‘healthy’	**afoʔo**	‘hole’	**tuf-**	‘to spit’
Kam	**fayya**	‘healthy’	**afoo**	‘mouth’	**tuf-**	‘to spit’
Sid	**fayyaaleessa**	‘healthy’	**ʔafo**	‘mouth’	**tuf-**	‘to spit’
Ged	**fayya**	‘healthy’	**afaʔo**	‘mouth’	**tuf-**	‘to spit’

Further evidence for PHEC ***f** comes from PHEC [339] **ful-* 'to go out, exit', [415] **afal-* 'liver', [47] **k'uf-* 'to cough', and PNHEC [65] **faras-* 'horse'.

Unlike the labial stop ***b**, there are no reconstructable instances of geminate ***ff** in HEC. The few cases of **ff** that do appear are clearly onomatopoetic in nature (cf. Kambaata *uff a?-* and Sidaama *uffu ass-* 'blow (on fire)').

There is limited evidence for PHEC ***f**. The majority of the instances of the fricative **f** are found in loanwords in the HEC languages, and it appears more frequently in Burji. There are instances of non-initial PHEC ***f** developing from an earlier ***b** discussed in §3.3.1.1. Another change ***s** → **f** showing Oromo influence on Burji is discussed in §2.3.1.2, PHEC **s*.

Summary

The labial consonant series (***b**, ***m**, ***f**) is securely reconstructed for PHEC, with systematic correspondences across all five HEC languages supporting each proto-phoneme. The examination of labial consonants in HEC reveals systematic patterns in both their historical development and synchronic distribution. While there is remarkable stability across all labials in initial position, non-initial environments trigger various developments, most notably in ***b**, which undergoes spirantization, lenition, and eventual loss in intervocalic position. The nasal ***m** demonstrates greater stability, showing only limited alternations through partial assimilation before alveolar-initial suffixes, while also serving as a crucial morphological marker in the passive suffix **-am**. The fricative ***f**, though it does occur in geminate form, exhibits regular correspondences across the languages, with some instances developing from earlier ***b** through conditioned changes at morpheme boundary. These developments align with broader tendencies in Cushitic historical phonology, particularly in the preference for lenition in intervocalic positions and the stability of initial consonants, while also revealing innovations specific to the HEC languages.

2.1.4 Glides

The glides ***w** and ***y** pose distinctive reconstruction challenges. While their reflexes in initial position show regular correspondences, intervocalic positions show extensive loss, making it difficult to distinguish inherited glides from epenthetic segments that emerge to resolve hiatus (§§3.4.1.2, 3.3.4).[45]

[45] Both ***w** and ***y** serve important morphophonological functions in the HEC languages, appearing in derivational processes (§4.1.1.2, §4.1.5.2, §4.2.5, §4.2.4.1).

This section examines the historical development of these glides and their distinctive patterning across the HEC languages, with particular attention to positional constraints and their interaction with surrounding vowels.

Glide rules

(a) Loss of intervocalic glides
(b) Gemination rule of intervocalic glides

PHEC *w

The following sets show the reflexes of the securely reconstructed PHEC ***w**:

	[293]		**[304]**		**[252]**	
PHEC	***waʔ-**	‘water’	***wit’-**	‘to sow, scatter’	***hasaaw-**	‘speak’
Burji	**waa**	‘water’	**wat’ar-**	‘scatter, spread’	**hasaaw-**	‘to talk’
PNHEC	***waʔ-**	‘water’	***wit’-**	‘to sow, scatter’	***hasaaw-**	‘speak’
Had	**woʔo**	‘water’	**wit’-**	‘to sow, to seed’	**hasaaw-**	‘speak’
Kam	**woʔa**	‘water’	**wit’-**	‘sow, scatter’	**haasaaww-**	‘discuss’
Sid	**wa**	‘water’	**wit’-**	‘sow’	**hasaaw-**	‘speak’
Ged	**wedeʔ-**	‘water’	**wiisa**	‘sow’	**haasoʔ**	‘speak’

In [293], rounding of ***a** → **o** after **w** is plausible in Hadiyya and Kambaata. However, all Gedeo examples are problematic: [293] *wedeʔ-e* and [304] *wiisa* have unique **d** and **s**, respectively,[46] while [252] *haasoʔ-* has lost the non-initial **w** altogether. Other reconstructable series with non-initial ***w** are difficult to find, and none exist for ***w** in intervocalic position.

More examples of PHEC ***w** include PHEC [139] **wozan-* → PNHEC **wozan-* ‘heart’, PHEC [259] **wol-* → PNHEC **wol-* ‘(each) other’, PHEC [309] **war-* → PNHEC **war-* ‘to gossip, tell the news’, PHEC [440] **weeʔn-* → PNHEC **weeʔn-*, **weeʔn-i-ʃʃ-* (sgv) ‘Colobus monkey’, PNHEC [406] **wees-* ‘ensete tree (Ensete ventricosum)’, and PHEC [252] **hasaaw-* → PNHEC **hasaaw-* ‘speak, talk’, and PHEC [443] **uħw-* → PNHEC **uħw-* ‘to give’.[47]

[46] See Sheko *yiits’* ‘to sprinkle’.
[47] See PEC **-ħiw-/*-ħuw-* (or **-hiw-/*-huw-*) → PHEC **ʔuhw-* ‘give’ (Sasse, 1982, p. 185).

The intervocalic glide loss ***w** → **ø** rule ((iii)a) may represent a further development of the lenition process ***b** → ***w** rule (ii), though the chronological relationship remains uncertain.

iii. Loss of intervocalic glides
   a. ***w** → **ø** /**V_V**
   b. ***y** → **ø** /**V_V**

The scarcity of evidence for non-initial ***w** and the lack of any single intervocalic ***w** indicate that this must be an old rule (attested in East-Omo-Tana as well).

There are limited instances of non-initial **w** in clusters and geminates.

wʔ: PHEC [226] **c'iwʔ-*, **c'awʔ-* 'chick(en)' → Burji *c'iwʔe*; PNHEC **c'aac'urr-*, **c'iic'oorr-* via **c'ac'~c'u-* (**c'iu-e*) → Ged *c'aac'ute*, Sid *c'aac'c'urre*, and via **c'ic'~c'o-rr-* (pl) → Had *c'iic'oolla*, *c'iic'oora*, Kam *c'iic'oorra*.

***wm**: PHEC [399] **dawm-* 'red' → Burji *duwwaaga* (m), *duwwaada* (f); PNHEC **dawam-* 'red' → Had *daamaʔa* 'dark red', Kam *daama* 'specific colour of a cow (red and black, e.g. front black, back red)', Sid *duume* 'red'.

wh**/wk**: PHEC [466] **diwh-*, **diwk-* 'empty' → Burji *duwwa* 'empty and deserted house'; PNHEC **diwh-*, **diwk-* 'empty' → Had *diiha* 'empty', Kam *diiha* 'empty', Ged *duwwi* 'empty, bare, naked'.

The sequence development of the examples is as follows:

PHEC ***iwʔ**/***awʔ**, ***awm**, ***wh**/***wk** →	Burji **iwʔ-**, **uww, uww**
	PNHEC ***ii**/***aa**, ***awam**, ***iwh**
	→ Had and Kam **ii**, **aa**, **ii**
	Ged and Sid **aa**, Ged **uww**, Sid **uu**

In [399] and [466], Gedeo has *diimmo* 'red' and *duwwi*, *duwwika* 'empty, bare, naked' and Burji has *duwwa* 'empty and deserted house', which may reflect influence from Oromo *diimaa* 'red' and *duwwaa* 'empty'.

Gemination of some single intervocalic glides (in Burji (Tesfaye Baye, 2015, p. 43), Sidaama, and Gedeo) is provisionally proposed. However, there are more instances of single ***y** gemination than ***w**.

***w** → **ww** /**V_V** (rare)

***y** → **yy** /**V_V**

The ***w** → **ww** /**V_V** is possibly marginal or morphophonological in some HEC languages, such as Sidaama, Gedeo, and Burji. It can phonotactically block the loss rule of single intervocalic ***w** (iii). That is mainly evident in the plurative suffix **-uw-a*, which has been retained in Hadiyya, Sidaama, and Gedeo as *-u(w)wa* (see §4.1.1.2, Plurative *-uw-a).

Single intervocalic **w** is geminated in the Amharic word *billawa* 'knife' in Hadiyya and Kambaata [96] *billawwa,* but merges with **b** and is retained as single in Sidaama *billaba* (cf. Oromo, which has **ø** in *billaa* 'knife'). Kambaata [147] *hawwarru-ta* 'evening' is from a weakened ***b** of PNHEC **habar-* 'evening'. Consequently, such **ww** geminates are not original (especially found in the plurative suffix ***-aw-a** → ***-aww-a**). Intervocalic gemination affects **y** more frequently than **w**, with numerous attested examples for **y** but limited evidence for **w**.

There are instances of Burji **h** corresponding to NHEC **w** in initial positions, e.g., Burji [520] *huuwʔ-* 'to fill, to fill up' and Sidaama *woʔm-* 'all, every, full', with the **w** as a reflex of ***ħ** before ***u** in NHEC.

PHEC *y

Unlike the PHEC glide ***w**, the examples of initial ***y** are limited.

	[388]		**[28]**		**[284]**	
PHEC	***y**	'to say, do'	***ay-**	'who?'	***ayyaan-**	'festival'
Burji	**iy-, i-**	'say, do'	**ayyee**	'who?'	**ayyaanaa**	'festival'
PHEC	***y**	'to say'	***ay-**	'who?'	***ayyaan-**	'festival'
Had	**y-**	'say'	**ayye**	'who?'	**ayyaana**	'festival'
Kam	**y-**	'say'	**aye**	'who?'	**ayyaana**	'festival'
Sid	**y-**	'say, judge'	**aye**	'who?'	**ʔayyaana**	'festival'
Ged	**hiyy-**	'say, try, do'	**ayye**	'who?'	**ayyaana**	'festival'

PHEC [388] **y* 'to say, do' becomes Gedeo *hiyy-* with initial **h-** and is similar to Burji where **-i** insertion has occurred. Section 4.2.4.1 includes a detailed

discussion on this (support) verb within HEC.[48] In PHEC [28] **ay-* 'who?', the geminated **yy** of Burji, Gedeo, and Hadiyya is likely due to an epenthetic **-y-**, while Sidaama and Kambaata *aye* (← **a-y-e*) have inserted **-y-** after the loss of intervocalic single ***y** via rule (ii(b)), all after the suffixation of the vocalic morpheme.[49] Eyob Kelemework (2015, p. 107) states that "[t]he nominative form of the pronoun 'who' is [ayye]; the degeminated form [aye] and the short form aa are also used as an interrogative pronoun" in Gedeo. There are similar cognates with geminate **yy** and single **y** in the rest of EC. Finally, PHEC [284] **ayyaan-* 'festival, holiday' has Cushitic cognates but is possibly a loanword. It has parallels in Ethiosemitic (Sasse, 1982, p. 30).

Further evidence for PHEC ***y** includes PHEC [749] **yaaʔ-* 'assembly, meeting', [167] **fayy-* 'healthy', [91] **mooyy-* 'mortar', and [448] **gaayy-* 'pipe (of clay)' (cf. Hadiyya *gaawwa* 'water pipe' and Amharic *gayya* 'pipe').

Evidence for initial PHEC ***y** is scarce. In addition to the above, there are PHEC [354] **yaaʔ-* 'to flow, assemble' and [274] **yeyy-* 'wolf, jackal', which is found in Burji *yeyyee* and Gedeo *yaayye*, while Sidaama has inserted initial **h-** in *hiyy-eessa*.

Similar to the PHEC glide ***w**, single non-initial ***y** is vulnerable in HEC. There is no evidence of a single intervocalic ***y** within the root in HEC, in accordance with the rule of intervocalic glides loss ((iii)b) ***y** → **ø** /**V_V**.

Such loss of single intervocalic **y** has been reported, for example, in Burji (Tesfaye Baye, 2015, p. 43) and Sidaama (Yri & Pepper, 2019). In Gedeo, there are cases of elision of stem-final **y** and **w** (Eyob Kelemework, 2015, p. 87).

The glide gemination of originally non-geminates (e.g., plurative suffix ***-uw-a** → ***-uwwa**) is likely a loss-blocking strategy in these languages (for Hadiyya, see Tadesse Sibamo, 2015, p. 23). Thus, besides the gemination of intervocalic ***w** discussed above, a similar one is tentatively posited ***y** → **yy** /**V_V**.

[48] Sasse (1982, pp. 102, 108) reconstructs it as a PEC stative prefix.
[49] The following derived forms could be proposed for Hadiyya, Kambaata, and Sidaama: **ay-y-e* (acc), **ay-y-i* (nom) 'who?'.

The following example shows no **y** reflex in Kambaata and Gedeo because they used a different support verb ***ʔ** (§4.2.4.2), and Gedeo **d'** is part of the reflexive denominative (§4.2.5.5).

> PHEC [31] **oo y-, *oo ʔ* 'to cry' → Burji *oy-* 'cry, weep'; PNHEC **oo y-, *oo ʔ* 'to cry' → Had *ooy-* 'to weep, mourn', Kam *oʔ-* 'cry, weep, mourn', Sid *ooy-*, *ʔoʔʔo yi*; *oʔo y-* 'cry', Ged *ood'-* 'wail, weep, mourn' (cf. Oro *oo(y)-* 'cry').

As Sasse (1982, p. 156) notes, this stem is a compound from PEC **ʔoo y-* 'say oo'.

Epenthetic **-y-** across morpheme boundaries is well attested in HEC. It is also well-attested in other EC languages, at least in EOT. A discussion and more details can be found under §3.4.1.2.

The following three forms show heterogeneous correspondences in which Burji **VyC** corresponds to **VVC** elsewhere in HEC:

> Burji *laymi*: PHEC [634] **leem-* 'bamboo' → PNHEC **leem-*, **leem-n-* 'bamboo' → Had and Kam *leema*, Sid, Ged *leemma* (cf. Oro *leemana* 'reed, bamboo').
>
> Burji *dureyʃi*, *dur-ee-ʃ*: PHEC [428] **dur-* 'rich' → Kam *duuballaaʃʃa*, Sid, Ged *dur-eessa*.
>
> Burji *ogeyʃi*: PHEC [236] **og-*, **og ʔ-y-ʃ-* 'traditional doctor' → PNHEC **og-*, **og-aʔ-y-ʃ* 'traditional doctor' → Had *ogeeʃʃa*, Sid *ʔogeessa*, *ʔogeette*, Ged *ogeeyyicca* 'wise man', *ogeeyitte* 'wise woman', *ogeeyy* 'wise men' (cf. Oro *ogeessa* 'wise, expert'.

Example [428] is likely Oromo-related (see Oromo *dureessa* 'rich', with the suffix **-ssa**; Lamberti & Sottile, 1997, p. 348),[50] [236] already contains derivational **-y-**, and only [634] is a lexical case. Since no trigger for root-internal epenthetic **-y-** is available in [634], I treat Burji *laymi* as an isolated

[50] Compare also Had *lik'-aayyis-* 'lend': PHEC [392] **lik'-* 'to lend' → Burji *lik'-ees-*; PNHEC **lik'-* 'to lend' → Kam *leʔ-ees-* 'lend of what cannot be returned; lend (money)', Sid *lik'-iiss-* 'lend ', Ged *lik'-eess-* (cf. Oro *lik'-eessa* 'lend (money)'). This shows that comparable glide-vowel sequences may also arise secondarily and are therefore not by themselves probative for reconstruction.

irregular development rather than as evidence for reconstructing proto ***VyC**.sequence (see §2.2.1.2 (PHEC *ee)).

The glides **w** and **y** are found in number-marking suffixes and occur epenthetically in various HEC languages in various HEC languages (see §4.1.1.2 and §3.4.1.2).

Summary

The limited evidence for non-initial ***w** and the systematic loss of intervocalic glides suggest these sounds were already unstable in PHEC. The contrast between relatively frequent **y** gemination and rare **w** gemination may reflect phonotactic preferences that differentiated the two glides beyond their basic articulatory properties.

2.1.5 Velars, Glottals, and the Pharyngeal

The velars, glottals, and pharyngeal consonants of PHEC show a distinct pattern of development across the HEC languages. This section examines the six consonants in this class: the velars ***g**, ***k**, and ***k'**, the glottals ***h** and ***ʔ**, and the pharyngeal ***ħ**. These sounds display both stable retentions and innovative developments in the descendant languages. The reconstruction of these consonants reveals important insights into the phonological structure of PHEC and the subsequent sound changes that have affected the individual HEC languages. Particular attention is given to the patterning of these consonants in different phonological environments and their participation in morpheme boundary alternation processes (§3.2, §3.3, §3.3.4).

2.1.5.1 Velars

Highland East Cushitic has three velars: the voiced **g**, the voiceless **k**, and the ejective **k'**.

PHEC *g

The regular **g** reflexes across all HEC languages in both initial and non-initial position support the secure reconstruction of PHEC ***g**. Burji, however, exhibits limited cases of **g** palatalization to **j** before **i** (§3.2.1).

	[36]		[702]		[80]	
PHEC	***gam-**	'to bite'	***hig-**	'to return, pass'	***gal**	'to enter'
Burji	**gam-**	'to bite'	**hig-**	'exchange'	**gal-**	'enter'
PNHEC	***gam-ʔ-** → ***gaʔm-**	'to bite'	***hig-**	'to return, pass'	***gal-**	'to enter'
Had	**gaʔm-**	'to bite'	**hig-**	'pass'	**gar-**	'pass the night'
Kam	**gaʔm-**	'to bite'	**hig-**	'pass'	**gal-**	'pass the night'
Sid	**gaʔm-**	'to bite'	**hig-**	'return'	**gal-**	'enter, pass the night'
Ged	**gaʔm-**	'to bite'	**hig-**	'return'	**gal-**	'pass the night'

Additional evidence for PHEC ***g** includes: PHEC [350] **sang-* 'ox, bull',[51] PHEC [126] **gog-* 'skin, hide', PHEC [67] **ang-* → PNHEC **ang-* 'hand, arm', PHEC [707] **magan-* 'sky, God', PHEC [223] **eg-* 'to wait', PHEC [85] **arg-* 'to send', and PHEC [438] **zuug-* 'to scrape, tan (hide)'.

Burji has limited instances of palatalized **j** corresponding to **g** in the rest of HEC before the high front vowel **i**. Some of them have a palatalized parallel in Oromo, where the palatalization is regular.

[400] Burji *jiira* 'fire': Kam *giira-ta* 'fire', Sid, Had, Ged *giira* 'fire'.[52]

[701] Burji *jisk-*, *jig-* 'to demolish/destroy': Had *giggiss-* 'to shake off', Kam *giggiʃʃ-* 'shake off, let fall down', Sid *giggiʃʃ-* 'tear down part of a house in order to renew it, snatch, kick off', Ged *giggis-* 'stutter'.

[183] Burji *jibb-* 'to hate, neglect': Had *gib-* 'flee', Kam, Sid, Ged *gib-* 'refuse', Oro *jibba* 'dislike'.

It is likely a contact-induced change applied irregularly in Burji, because [400], [701], and possibly [183] are the only three established examples that support the positing of a rule. Oromo [183] *jibba* 'dislike' with **j** is not the only such parallel. Here, there is a chance that it is borrowed into Burji. Clearer loans with **j** into Burji and other HEC languages include:

[640] Oro *jabaa* 'strong', *jabeeŋŋa* 'strength', *jabeessa* 'make strong': Burji *jebeenna*, *jebuntaa* 'strength': Sid *jawaante* 'strength',

[51] As Sasse (1982, p. 163) notes, [350] is a widespread word, but its origin is difficult to determine (see Amharic *sänga* and Oromo *sang-aa* 'ox for beef').
[52] Probably related to [940] Burji *jirmalleyi* 'acacia' and Had *giraar* 'acacia tree, mimosa'.

Ged *jabeessa* ‘strengthen’ (cf. Arbore *goob-* ‘become a strong young adult’, Amharic *jäbdu* ‘act of bravery’).

[265] Oro *hujj-ii* ‘work’: Burji *hujee* ‘work’: Kam *huji-ta* ‘work’, Ged *huja* ‘work’.

There are isolated instances of palatalized **j** in non-initial position [18], sometimes alongside the non-palatalized variant [708], discussed in §3.2.1. However, the regularity of **g** palatalization in Burji is further challenged by the fact that the only example that goes as far as PEC [4] shows non-palatalized **g** before **i** in Burji, although Oromo (and EOT languages) exhibit palatalized **j**.

[4] Burji *gilba* ‘knee’: Had *gurubbo* ‘knee’, Kam *gulubi-ta* ‘knee’, Sid *gulube* ‘knee’, Ged *gulubo* ‘knee’, Oro *jilba* ‘knee’.

[632] Burji *giddee* ‘inside’: Had *guda* ‘middle, half’, Sid *giddo* ‘inside’, Ged *giddo* ‘inside, in’, Oro *jidd-uu* ‘middle’, *goda* ‘inside, *gadi* ‘interior’.

These examples point to an irregular Burji-specific palatalization of ***g** → **j** before high front vowels, particularly **i**. Since the NHEC cognates consistently retain **g**, this change must post-date the split of Burji from PNHEC. Oromo is the most plausible contact source, because comparable palatalization is regular there, and Burji also shows clear Oromo loans with **j** in this section. The wider EOT parallels also provide a typological support for this established development in LEC.

While the voiced velar ***g** shows contact-induced instability in Burji, the voiceless velar ***k** exhibits a different pattern of distribution, particularly in non-initial position.

PHEC *k

PHEC ***k** is continued regularly in initial position. However, it mostly occurs as a geminate or in a cluster in post-vocalic position. Here are examples of the securely reconstructed PHEC ***k** reflexes in initial position:

	[705]		[418]		[30]	
PHEC	***kul-**	'to tell'	***ku, *ko**	'this' (m.nom, acc)	***kaʔ-**	'to get up, arise'
Burji	**kulʔud**	'to stutter'	**ku, ka**	'this' (m.nom, acc)	**kaʔ(ʔ)-**	'to get up'
PNHEC	***kul-**	'to tell'	***ku, *ko**	'this' (m.nom, acc)	***kaʔ-, *keʔ-**	'to get up, arise'
Had	**kur-**	'tell'	**kuk, ka**	'this' (m.nom, acc)	**kiʔ-**	'to get up'
Kam	**kul-**	'tell'	**ku, ka**	'this' (m.nom, acc)	**keʔ-**	'to get up'
Sid	**kul-**	'tell'	**kunni-, kónne**	'this' (m.nom, acc)	**kaʔ-**	'to get up'
Ged	**kul-**	'tell'	**kunni, konne**	'this' (m.nom, acc)	**kaʔa, keʔa**	'to get up'

Two forms [30] **kaʔ-* and **keʔ-* 'to get up, arise' are reconstructed for PNHEC because, in addition to vowel ablaut, the following glottal stop also causes variation like that seen in the three different vowel reflexes **a**, **i**, and **e** (see §2.2.2.1).[53]

Additional examples of initial **k** include: PHEC [652] **kiil-* 'to do magic', [478] **kott-* 'hoof, foot', [327] **kor-* 'saddle', [391] **kum-* 'thousand', PNHEC [670] **kaas-* 'to plant', and [602] **kin-* 'stone, rock'.

There is no evidence for a fully retained single non-initial ***k** (see Kambaata (Treis, 2008, p. 26)). It survives only in clusters (mostly **nk**, **rk**) and as a geminate (**kk**) in non-initial position.

> PHEC [51] **lokk-*, **lekk-* 'foot, leg' → Burji *lukka*; PNHEC **lokk-*, **lekk-* 'foot, leg' → Had *lokko*, Kam *lokka-ta*, Sid *lekka*, Ged *lekka* (see Oro *luka* 'foot, leg').
>
> PHEC [161] **bark-* 'head-support, pillow (of wood)' → Burji *bórk-e*; PNHEC **bark-* 'head-support, pillow (of wood)' → Had *barkuma*, Kam *borkaanu*, Sid *barko*, Ged *barko* (cf. Oro *bark-oo* 'headrest').
>
> PHEC [69] **midk-* 'right(-side)' → Burji *mirga* 'right(-side)'; PNHEC **midk-* 'right(-side)' → Had *makka*, Kam *makki-ta*, Ged *midda*.
>
> PNHEC [397] **kikkil-* 'armpit' → Had *kikkiro*, *kilkilo*, Kam *kikkili-ta*, *kilkilli-ta*; Sid *kikkiʔla*, *kikkilicco*.

[53] The origin of the glottal stop is a pharyngeal, which causes vowels to lower in some Cushitic languages. However, this is not the case in HEC, since the PHEC reflex of the pharyngeal ***ʕ** is the glottal stop ***ʔ**.

In [69], the **d** has assimilated to the following **k** in Hadiyya and Kambaata **madk-* with vowel assimilation/ablaut and to the preceding **d** in Gedeo via **mid-g-*. Oromo *mirga* ‘right(-side)’ likely influenced the Burji word with **r** and **g**. In the other examples [397], the root is reduplicated partially or fully in each language, resulting in single **k** and geminate **kk**.

There are morphophonological processes that contributed to the lenition and subsequent loss of potentially retained single non-initial **k** discussed in §3.2.2 and occasional palatalization in Burji §3.2.1.

PHEC *k’

PHEC ***k’** is only arguably reconstructable, and its historical interpretation remains unresolved. Below is a summary of the attested correspondence series covered in the analysis and their frequency:

	Burji	Had	Kam	Sid	Ged	Frequency
1.	**k’**	**k’**	**k’**	**k’**	**k’**	Common
2.	**k’**	**ø**	**ø**	**ø**	**ø**	Less Common
3.	**k’/k**	**k**	**k/g**	**g**	**k’**	Uncommon

Note that these series are chiefly attested in initial position, and the reflexes are mostly sporadic in the second and third series. At present, the evidence does not support a single regular account of all three series. Series 1 favors reconstructing PHEC ***k’**, whereas series 2 and 3 may reflect differential loss, analogical reshaping, or an earlier conditioning that is no longer recoverable from the surviving forms. As with the other non-pulmonics discussed in §2.1.2, the discussion below evaluates both the HEC distribution and the wider EC context.

In NHEC, inherited ***k’** is much less stable than Burji alone would suggest. Many NHEC words with **k’** lack corresponding non-pulmonic cognates elsewhere in EC and are therefore difficult to project back beyond HEC with confidence. Some may reflect reestablished PNHEC ***k’**, including early proto-level borrowings or post-split innovations (see §2.1.1.2 (PHEC *ʃ)); others may continue older PHEC ***k’** that was later lost outside NHEC or elsewhere in EC. For that reason, the relation between PHEC ***k’** and PNHEC ***k’** is better treated as unresolved than reduced to a single history.

	[189]		[392]		[108]	
PHEC	***k'araar-**	'be bitter'	***lik'-**	'to lend'	***sunk'-**	'to kiss'
Burji	**k'araar-**	'be bitter'	**lik'-ees-**	'to lend'	**sunk'-**	'to kiss'
PNHEC	***k'araar-**	'be bitter'	***lik'-**	'to lend'	***sunk'**	'to kiss'
Had	**k'araar-**	'be bitter'	**lik'-aayyis-**	'to lend'	**sunk'-**	'to kiss'
Kam	**k'araar-**	'be bitter'	**leʔ-ees-**	'to lend'	**sunk'-**	'to kiss'
Sid	**k'araar-**	'be bitter'	**lik'-iiss-**	'to lend'	**sunk'-**	'to kiss'
Ged	**k'araar-**	'be bitter'	**lik'-eess-**	'to lend'	**sunk'-**	'to kiss'

Two sets stand out because they preserve initial **k'** in NHEC together with broader EC support. One is [189], which occurs with **k'** in Dahalo *k'araare* 'bitter' and **q** in Somali *qaraar* and Maay *qereer* 'bitter'. The other is [115], where Gedeo, Hadiyya, and Sidaama *k'iid-* 'become cold' has multiple EC parallels with initial **k'** or **q** as well. In [392], by contrast, Kambaata shows a puzzling **ʔ** reflex instead of **k'**. This is not an isolated anomaly: **k'** is elsewhere replaced by **ʔ** or **ø** in Kambaata and the other NHEC languages. At least some of these losses may therefore reflect an earlier conditioning that is no longer recoverable from the surviving lexicon.

For more examples of PHEC ***k'**** see PHEC [675] **k'oom-* 'calabash', [449] **k'al-* 'give birth, beget (of animals)', [582] **hak'k'-* 'wood', [240] **k'ort'umm-* 'fish' (← Oromo), PHEC [116] **k'ok'k'-* 'throat' → PNHEC **k'oonk'-* 'throat', and PNHEC [494] **k'ob-, *k'ok'~k'obb-* 'to surround, to bend', [49] **k'uub-* 'egg', PNHEC [21] **c'ak'as-* 'to listen, spy'.

The non-pulmonic velar ***k'**** discussed here contrasts with glottalized velars arising through fusion. In PNHEC, sequences of velar obstruent plus glottal stop undergo fusion to yield geminate **k'k'** (§3.3.1.3). These glottalized consonants appear in derived forms where the glottal stop belongs to a suffix or stem-final or additional suffix segment, whereas non-pulmonic ***k'**** occurs in stem-initial and non-initial positions independently of morphological structure.

The examples below show initial **k'** in Burji corresponding to **ø** in NHEC. In contrast to the above examples, most of the following have parallels in the rest of EC, where **k'** and other correspondences are also attested. Therefore, at least some of them are arguable retentions in Burji.

[29]	k’	ø	Gloss
PHEC	***k’ab-**		‘to hold/seize’
Burji	**k’af-**		‘to have’
PNHEC		***ab-**	‘to hold/seize’
Had		**amad-**	‘hold, seize’
Kam		**ap’p’-**	‘catch fire, hold, seize’
Sid		**amad-**	‘hold, seize’
Ged		**abid-**	‘hold, seize’
cf. Oro	**k’aba**		‘have, possess, seize’

In [29], Burji and Oromo **k’** corresponds to **k’** in numerous EC languages:

> Dir *k’ap-* ‘catch, have’, Mos *k’aɓa* ‘catch’, Boni Kilii *k’aba?-* ‘to catch, seize’, Arb *k’abad’-* ‘seize’, and Gaw *ɠap-* ‘catch, seize’.
>
> Saho *ab-it* ‘take for oneself’, Afar *-ibbiɗ-* ‘catch’; Bay *abat-* ‘catch, seize personally’, Boni Bireeri *owd’-*, Boni Jara *obəd’-* ‘to catch, seize’, Elm *ap* ‘to have, to seize’.

The initial **ø** in PNHEC **ab-* ‘to hold/seize’ also has **ø** correspondences in Saho, Afar, Bayso, Boni, and Elmolo.

[74]	k’	k, g	ø	Gloss
PHEC	***k’ub-**	***-kub-**		‘to fall (down)’
Burji	**(k’up’an-)**	**(gambun?-)**		‘to squat, to fall by itself on the face’
PNHEC			***ub-**	‘to fall (down)’
Had			**ub-**	‘fall (down)’
Kam			**ub-**	‘fall (down)’
Sid			**ub-**	‘fall (down)’
Ged			**ib-**	‘fall (down)’
cf. Oro		**kufa**		‘fall (down)’

In [74], the two reconstructions **k’ub-* and **-kub-* ‘to fall (down)’ remain competing possibilities. If one accepts the EC-supported reconstruction **-kub-* ‘to fall (down)’, PNHEC **ub-* would require loss of the initial velar under a conditioning that is no longer directly observable.[54] If one instead privileges

[54] The EC parallels are Rendille, Somali, Boni *kuf* ‘to fall’, and Yaaku *kup-* ‘die’. One assumption may be a remnant prefix verb root preceded by a vowel that triggered the lenition and elision of the **k** (§3.2.2).

the Burji form, the set joins the broader **k’ ~ ø** series but remains irregular. The set, therefore, does not resolve the history of ***k’**; rather, it shows that the correspondence cannot yet be reduced to a single regular development.

The correspondence of non-pulmonic **k’** with the pulmonic **k** is not an isolated case, as discussed below.

[71]	**k’**	**ø**	**Gloss**
PHEC	***k’utaħ-**		‘thorn’
Burji	**k’uwwa, k’uwa**		‘thorn’
PNHEC		***utaħ-**	‘thorn’
Had		**utta**	‘thorn’
Kam		**uta**	‘thorn’
Sid		**uta**	‘thorn’
Ged		**ute**	‘thorn’
cf. Oro	**k’orraattii**		‘thorn’

The EC parallels of [71] with initial **k’/q** include:

> Dir *k’uɗɗ-éét* ‘thorn, words which prick’, Mos *k’uɗɗeca*, Som *qodaħ*, Maay *quruŋ*, Arb *k’ed’eh*, Ts’am *q’ants’-e* ‘thorn’.

Various other correspondences are found in Konso *ʃaattalʃaʔatta*, Saho *keena* ‘thorn, prickly’, Afar *keena*, Rendille *kuɖáʜ*, and Boni *uŋa(h)* ‘thorn’.

[381]	**k’**	**ø**	**Gloss**
PHEC	***k’ul-, *k’ul~k’ul~l-ʔ-**		‘leech’
Burji	**k’ulaallee**		‘leech’
PNHEC		***ul-, *ul~ul~l-ʔ**	‘leech’
Had		**urulla**	‘leech’
Kam		**urulla**	‘leech’[55]
Sid		**ulaaʔulle**	‘leech’
Ged		**ulaʔulla**	‘leech’
cf. Oro		**ulaanʔul-a**	‘leech’

In addition to Oromo *ulaanʔul-a* ‘leech’, the only available EC parallels of [381] are Afar *ʕallullé*, Bayso *halalac’e*, and Somali *ʕalaaʕul* ‘leech’, with none

[55] The Kambaata **r** reflex of ***l** is probably due to Hadiyya influence, where such a change is regular.

supporting the reconstruction of PHEC **k’**. Even so, the PHEC reconstruction is retained only provisionally. Burji **k’** places these sets within the broader initial **k’ ~ ø** pattern, but the wider EC evidence is too uneven to decide whether the Burji forms are genuine retentions, secondary preservations under contact influence, or mixed cases. The corresponding Oromo **k’** in two of the four sets may reflect contact influence, but it does not by itself determine directionality. What is clearer is that PNHEC shows initial **ø**. Accordingly, the PNHEC reconstructions are kept with **ø**, while the PHEC reconstructions with ***k’** are treated as provisional and unresolved. Some may reflect genuine PHEC retentions later lost elsewhere in HEC or EC; others may not.

There is a marginal series of **k’**, **k**, and **g** correspondences (see [74] above) attested in the following few examples:

[47]	**k’**	**k**	**Gloss**
PHEC	***k’uf-**		‘to cough’
Burji	**k’ufey-**		‘to cough’
PNHEC		***kuf-**	‘to cough’
Had		**kut’t’-**	‘cough’
Kam		**kuc’c’-**	‘cough’
Ged	**k’ufa-aʔ-**		‘cough’
cf. Oro	**k’ufaʔa**		‘cough’

Here are the EC parallels in set [47]:

> Dir *k’ayy-oodˀ-*, Kon *k’aaw-adˀ-*, Som *qufaʕ*, Maay *quffi* ‘cough’, Boni *k’ufaʔ-*, Das *gˀuf-ume*, Yaaku *qopeʔe-*, Ts’am *gˀufaʕ* ‘to cough’.
>
> Saho *ufuʕe* ‘to cough’, Bay *ufii-* ‘cough’, Boni *ʿufaʔ-*, Elm *úfe*, Arb *ʔuf(e)ʔ-*, Das *ʔúfuk* ‘to cough’.

PHEC **k’** corresponds to **k’**/**q**/**gˀ** in most of the EC parallels and, as usual, **ø** in Saho, Bayso, Boni, and Elmolo in addition to Arbore and Dasanach. Rather than **ø**, PNHEC shares **k** with Afar *kaʕite*, *kaʕùw-eħʕe* ‘cough’.

[115]	k’	k	g	Gloss
PHEC	***k’ab y-t-**			‘become cold’
Burji		**kabb-**		‘cool, become cold’
PNHEC	***k’iid-**			‘become cold’
Had	**k’iid-**			‘cool, become cold’
Kam			**gid-**	‘cool, become cold’
Sid	**k’iid-**			‘cool, become cold’
Ged	**k’iid-**			‘cool, become cold’
cf. Oro	**k’abb-an-**			‘become cool (liquid)’

As in the previous set [115], PHEC ***k’*** corresponds to EC **k’**/**q**:

> Dir *k’appann-aw-* ‘be cool (weather or personality (calm))’, Mos *k’appannayya*, *k’apana* ‘(be) cold’, Kon *k’ap-p-ann-aaw-* ‘become cool (liquid)’, Ren *xóbo* ‘coolness, cool’, Som *qabow* ‘coldness, cold’, Maay *qoboob* ‘cold’, Arb *k’abbat* ‘cold (thing)’.
>
> Bay *ambalki* ‘cold, low temperature’, Boni *abˈóob* ‘coolness, cool’; Das *ʔabbúnab* ‘cold’, Elm *ʔápat* ‘shadow; become cold’.

Even though PNHEC has ***k’***, Bayso, Boni, and Elmolo have **ø** as usual.

[288]	k’	g	Gloss
PHEC	***k’el-**		‘snail’
Burji	**k’ilaank’illoo**	**k’ilangiloo**	‘snail’
PNHEC	***k’el-**		‘snail’
Had	**k’elema, k’eloʔo**		‘snail’
Kam	**k’eloma-ta**		‘snail’
Sid	**k’erek’enicco**	**gerk’eʔnâma**	‘snail’
Ged	**k’ak’ank’ale**		‘snail’

There are no solid EC parallels for [288] (see Oromo *c’iilalluu* ‘snail’).

The irregular reflex **g** is attested in Burji [288] *k’ilangiloo* ‘snail’, Sidaama *gerk’eʔnâma* ‘snail’, and Kambaata [115] *gid-* ‘cool, become cold’. Tesfaye Baye (2015, p. 79) notes that in Burji, “/k/ is sometimes found in free variation with /g/ based on dialectical differences. People from the northern area of Burji and the surrounding of Soyama use the variant /k/ while those from the southern use /g/…”. This explains the above **k’ ~ g** variation in [288]. Such variation has also been reported by Sasse (1982, p. 18).

The velar consonants ***g** and ***k** show relatively stable development across the HEC languages. By contrast, the non-pulmonic ***k'** remains an unresolved problem with a series supporting an arguable reconstruction of PHEC ***k'**, but two series that do not yet yield a single regular explanation. Some mixed reflexes may continue an older conditioning that is no longer recoverable; others may reflect differential loss, analogical reshaping, or contact. Likewise, PNHEC ***k'** cannot be assigned a single origin: some forms are likely reestablished within PNHEC, while others may preserve older PHEC lexicon lost elsewhere in EC. For the present, the most defensible conclusion is therefore that PHEC ***k'** is arguable but not resolved, and that the three-way contrast among PHEC ***g**, ***k**, and ***k'** is best treated as a provisional working reconstruction rather than a fully regularized historical result. The distributional patterns indicate that ***k** frequently occurs as a geminate or in consonant clusters in non-initial positions, while ***k'** shows a limited distribution in non-initial positions.[56] These observations align with broader tendencies in East Cushitic historical phonology, particularly the retention of initial consonant contrasts across the language family.

2.1.5.2 Glottals

The two HEC glottal consonants are **h** and **ʔ**. Elision or lack of non-initial glottal consonants (intervocalic, syllable coda, in a cluster, etc.) has been reported by Sasse (1982, pp. 18, 35) and Tesfaye Baye (2015, p. 61) for Burji, by Eyob Kelemework (2015, pp. 46, 83–86) for Gedeo, by Tadesse Sibamo (2015, p. 34) for Hadiyya, and by Anbessa Teferra (2000, p. 27) for Sidaama. Similar to the sonorants, the glottal consonants **h** and **ʔ** are also used for epenthesis in HEC (§3.4.2).

PHEC *h

Initial PHEC ***h** is continued in all HEC languages, so its reconstruction is secure.

[56] Irregular non-initial **k'(k')** should be evaluated against three possible sources: direct inheritance, contact-induced variation, and fusion as a result of boundary alternations. The fusion path is particularly relevant for forms showing geminate **k'k'** in PNHEC post-vocalically, as these systematically arise from velar-glottal sequences (§3.3.1.3) rather than continuation of inherited geminate **k'k'**.

	[228]		[702]		[380]	
PHEC	***harr-**	‘donkey’	***hig-**	‘to return, pass’	***hik’k’-, *hek’k’-**	‘have hiccups’
Burji	**harree**	‘donkey’	**hig-**	‘exchange’	**hirc’immeed’-**	‘have hiccups’
PNHEC	***harr-**	‘donkey’	***hig-**	‘to return, pass’	***hik’k’-, *hek’k’-**	‘have hiccups’
Had	**halla**	‘donkey’	**hig-**	‘pass’	**hik’k’aʔ-**	‘have hiccups’
Kam	**harri-ta**	‘donkey’	**hig-**	‘pass’	**hik’k’ aʔ-**	‘have hiccups’
Sid	**harre**	‘donkey’	**hig-**	‘return’	**hek’k’i yaa**	‘have hiccups’
Ged	**harre**	‘donkey’	**hig-**	‘return’	**hek’k’eʔ mef-at-**	‘have hiccups’

Set [380] also suggests Oromo influence. Its Oromo parallel, *hirk’inf-* ‘have hiccups’, contains **f**, whereas the corresponding HEC forms are irregular in this position: Burji has **r** (see §3.4.4) and Gedeo has **f**. Since Oromo **f** regularly reflects earlier ***s** → **f** (see §2.1.1.2 (PHEC *ʃ)), probably in a causative/derivational suffix in **-s**, the Gedeo **f** is best treated here as contact-aligned rather than as an inherited regular reflex.

Further examples of PHEC ***h** include: PHEC [492] **haar-* ‘new’, [256] **ham ʔ* ‘to slander’, [768] **hab-* ‘to forget’, [461] **has-* ‘to want, seek’, [493] **hiiʔ-* ‘take!’, PHEC [722] **hoog-* ‘to clean’, PNHEC [700] **haʔm-i-ʃʃ-* ‘ensete tree root/corm’, and [601] **hos-* ‘to pass/spend the day’.

Overall, reflexes of reconstructable **h** are mainly attested before the low vowel **a**, less commonly before the front vowels **e** and **i**, rarely before **o**, and lacking before the back vowel **u**.[57] The less common occurrence of **h** before vowels other than **a** might have contributed to an irregular analogical dropping of initial **h** before front vowels in some HEC languages, e.g., in Kambaata *iinc’aʔ-* ‘stretch (body)’ ← PHEC [390] **hiit’-* ‘to stretch’, *ĩ* ‘take!, here’ ← PHEC [493] **hiiʔ-* ‘take!’.

The HEC languages have largely lost non-initial ***h**. The relevant morphophonological details are in §3.2.2, where the lenited **h** from PHEC ***k** is discussed.

[57] There are instances of Burji **h** corresponding to NHEC **w,** which could be taken as evidence for reconstructing ***hu**. However, there is also room to argue for a merger with ***ħu** and a reconstruction of ***ħ** for such series.

PHEC *ʔ

In HEC, the glottal stop **ʔ** appears as a retention of ***ʔ** in monosyllabic stems and serves multiple morphological functions. It serves as a reflexive and genitive suffix, and is also found as an epenthetic sound. Below are reflexes of the securely reconstructable non-initial ***ʔ**, which are presumably part of the stem:

	[431]		[34]		[293]	
PHEC	***raʔ-**	'ripen, be ripe'	***baʔ-**	'be lost'	***waʔ-**	'water'
Burji	**raʔ-**	'ripen, be ripe'	**bey-**	'be lost'	**waa**	'water'
PNHEC	***raʔ-**	'ripen, be ripe'	***baʔ-**	'be lost'	***waʔ-**	'water'
Had	**liʔ-**	'ripen, be ripe'	**biʔ-**	'be lost'	**woʔo**	'water'
Kam	**reʔ-**	'ripen, be ripe'	**baʔ-**	'be lost'	**woʔa**	'water'
Sid	**leʔ-, raʔ-**	'ripen, be ripe'	**baʔ-**	'be lost'	**wa, waa**	'water'
Ged	**raʔ-**	'ripen, be ripe'	**baʔ-**	'be lost'	**wedeʔe**	'water'

As discussed in §2.2.2.1, the vowel variation before a glottal stop is found even within the same language in the above examples: Sidaama [431] *leʔ-*, *raʔ-* 'ripen, be ripe', and Burji [34] *bey-* (Roba Dame & Wedekind, 2008, p. 8), *be-*, *bay-* (Hudson, 1989, p. 180) 'be lost', and Gedeo [293] *wedeʔe* (Wedekind, 2008, p. 147), *wadaʔe* (Hudson, 1989, p. 165) 'water'.[58] The **d** and **e** in Gedeo [293] *wedeʔe* 'water' might be suffixes, while **ʔ** is epenthetic (§3.3.4.1). Although vowel variation following **w** is common, at least in NHEC, the glottal stop is the actual trigger of this vowel alternation, as demonstrated by the fact that this phenomenon includes Burji and all three examples (§2.2.2.1).

Post-vocalic **ʔ** in Burji has different correspondences conditioned by boundary alternation processes (§3.3.3). Burji [431] *raʔ-* 'ripen, be ripe' has a preconsonantal stem *raa-* (Sasse, 1982, p. 157) since **ʔ** is dropped before consonants with concomitant lengthening of the preceding vowel (Sasse, 1982, p. 18).[59] However, the causative stem of [34] Burji *bey-* 'be lost' is *ba-ss* (caus) (Sasse, 1982, p. 34) without vowel doubling, similar to *t'ay-* 'to extinguish' becoming *t'as-s-* in the causative but with the fusion of **ʔ** and **s** into **ss**. This means that vowel doubling is not regularly applied in Burji when **ʔ** is dropped.

[58] Burji also has *wáyu* (*wéu*) (subj), *way-**i**-n-* (gen) 'water' (Sasse, 1982, p. 186).
[59] Another example is Burji *raʔeed'aa(ga)* (adj.m), *raaʃeettaada* (adj.f) 'cooked, ripe'.

Intervocalic **ʔ** is scarcely found in Burji (Tesfaye Baye, 2015, p. 30).

> PHEC [144] **meʔ-* 'how many? how much?' → Burji *miʔa* 'how many? how much?'.
>
> PHEC [493] **hiiʔ-* 'take!' → Burji *heʔee* 'take!, here', *heettaa* (interj) 'Here it is!'.

Otherwise, non-initial **ʔ** in Burji is found as a suffix (§§4.2.2.1, 4.1.2) and following sonorant consonants, in contrast to the NHEC languages (§3.3.2). There are three regular morphophonological processes concerning **ʔ** in Burji, discussed in detail in the respective sections:

(a) elision with compensatory vowel doubling (§3.3.3),
(b) **-y-** epenthesis (§3.4.1.2),
(c) complete progressive assimilation (§3.1.1.3)

Therefore, in addition to fully retained **ʔ**, a glottal stop is reconstructed in non-initial position where **ʔ** is found in NHEC corresponding to compensatory vowel doubling, epenthetic **-y-**, or geminated suffix consonant in Burji. The evidence is almost exclusively found in monosyllabic stems. The instances of HEC glottal stop found in disyllabic or multisyllabic stems are mostly epenthetic and affixes.

The previous paragraphs examined the reflexes of PHEC ***ʔ** in non-initial position and determined their role as suffixes only. Initial **ʔ**, on the other hand, lacks consensus as to its phonemic status in HEC. Therefore, the remaining part of this section on PHEC ***ʔ** synthesizes the synchronic evidence documented for initial **ʔ** in the individual languages and argues why initial ***ʔ** cannot be reconstructed.

Most HEC grammars posit initial **ʔ** for vowel-initial roots, including Burji (Tesfaye Baye, 2015, pp. 47, 30, 31), Sidaama (Yri & Pepper, 2019), Hadiyya (Tadesse Sibamo, 2015, p. 33), and Kambaata (Treis, 2008, p. 26). Eyob Kelemework (2015, p. 41) postulates vowel-initial roots for Gedeo, conforming to Hudson (2007) but disagreeing with Wedekind (1985).[60]

Historically, more than one consonant (e.g., PEC ***ʕ** and ***k'**) has merged into **ø** in initial position in HEC. The pattern of merging and influence on the

[60] There are differences between orthography and phonemic transcription (see Yri & Pepper, 2019).

following vowel, if any, is unclear. Thus, vowel-initial roots are reconstructed in this study without a default glottal stop.

Further arguments in favor of vowel-initial stems include the existence of vowel-initial suffixes in Kambaata, Gedeo, and Hadiyya (e.g., case vowels). In Kambaata, there is no evidence of a glottal stop when vowel-initial nominals are reduplicated or in compounds (Treis, 2008, pp. 27, 178).

Stem(s)	**Gloss**	**Reduplicate/ Compound**	**Gloss**
ʔisso ʔóot	'they'	**ʔísisso ʔóont**	'they … each other'
min- + ʔanná	'house' + 'father'	**minanná**	'householder'
abb-á + am-á-ta	'big' + 'mother'	**abb-am-á-ta**	'grandmother'
lank-í + am-á-ta	'second' + 'mother'	**lankaam-á-ta**	'sister of mother'[61]

Contrary to the above analysis, Treis treats the initial glottal stop as a phoneme, as evident in the reduplicated initial syllable of *ʔontú* 'five' → *ʔóʔontú* 'five each' and she explains *ʔísisso ʔóont* 'they … each other', *ʔaʔn-aʔno ʔóot* (pl) 'you' (p. 358) as rather a reduplication of the first vowel and the subsequent consonant. In the compound *minanná* 'head of household', the explanation is that the stem-initial glottal stop of the second constituent is usually deleted. Those are three different analyses, so it might be convenient to instead posit that all of these stems begin with a vowel and assume epenthesis of **ʔ** in *ʔóʔontú* intervocalically (§3.3.4.1). Treis (2008, p. 358, fn. 34) acknowledges the lack of **ʔ** in the reduplicated *ʔísisso ʔóont* 'they … each other' and *minanná* 'head of household', saying they "might be considered an argument not to treat the glottal stop as a phoneme but as a default word-initial phone".

The vowel-initial stems are also evident in Gedeo's phrasal expressions.

Stem(s)	**Phrase**	**Gloss**
giddo ʔuggi	**giddougg**	'put it inside!'
ʔayyi -ʔa	**ʔayyia**	'whom-for?'
hoola ʔabbo	**hoolabbo**	'oh dear!'

The lack of an intervening consonant results in diphthongs (**ou** and **ia**) when the vowels are different, as in the first two examples, and elision of one vowel when they are identical, as in the third example (Wedekind, 1980, p. 161). Rather than taking the lack of the glottal stop in such forms as evidence for

[61] Additional examples include: *abb-á* 'big' + *ann-á* 'father' → *abb-ann-á* 'grandfather' and *lank-í* 'second' + *ann-á* 'father' → *lankaann-á* 'brother of father'.

vowel-initial stems, Wedekind describes the above examples as instances of initial glottal stop elision.

In Hadiyya, as well, the vowel-initial stems mostly show no glottal stop in their compound formations.[62]

Stem(s)	Gloss	Compound	Gloss
moóllo + ʔánna	'ethnic group' + 'father'	**moollánna**	'lord/god'
ʔallaácca + ʔuúlla	'keeping' + 'land'	**ʔallaccuúlla**	'pasturage'
ʔáma + ʔaáyya	'mother' + 'sister'	**ʔamaáyya**	'aunt'
ʔánga + ʔoóbba	'hand' +'palm'	**ʔangoóbba**	'palm'
máto + ʔílle	'one' + 'eye'	**matílle**	'one-eyed person'[63]

There are three Hadiyya examples with a glottal stop in compounds:

Stem(s)	Gloss	Compound	Gloss
míne + ʔáma	'house' + 'mother'	**miʔnáma**	'householder (mother)'
míne + ʔánna	'house' + 'father'	**miʔnánna**	'householder (father)'
ʔeeliínco + ʔaágga	'sun' + 'entrance'	**ʔeeliinciʔaágga**	'morning'

These examples probably involve complex genitive constructions (see Genitive glottal stop ʔ in §4.1.3.1).

In sum, despite the deficient distribution, the "/ʔ/ is a fully-fledged consonant with all the 'human rights' of any consonant" (Yri, 2004). Beyond its status as a retained proto-phoneme, PHEC ***ʔ*** plays a morphophonological role in PNHEC. The glottal stop triggers fusion when adjacent to obstruents, yielding geminate glottalized consonants (§3.3.1). This morphophonological function makes ***ʔ*** central to understanding the distribution of consonants in modern HEC languages, as inherited non-pulmonics should be distinguished from the glottalized geminates derived from fusion rather than from direct inheritance. Further discussion of the non-initial glottal stop in morpheme boundary alternations is detailed in §3.3 and in morphology in §4.1.4.1, §4.1.3, §4.2.3.1, §4.2.5.1 and §4.2.4.2.

---

[62] It is unclear whether these forms are compounds, as Tadesse Sibamo (2015) writes them as a single form, whereas Ritter (2007) presents them in two separate forms.

[63] Additional examples include: *lob-* 'great' + *ʔánga* 'hand' → *lobángo* 'helpful, valuable', *ʔánna* 'father' + *ʔáre* 'wife' → *ʔannáre* 'step-mother', and *ʔánna* 'father' + *ʔabbáyyo* 'brother' → *ʔannabbaájjo* 'uncle'. Data source: Tadesse Sibamo (2015).

2.1.5.3 The Pharyngeal

PHEC *ħ

The correspondence series Burji **h**/**ø** and NHEC **h**/**w**/**ø**, conditioned by adjacent vowels and word position, requires explanation. Two analyses compete: (1) an analysis as epenthesis, treating these segments as phonologically inserted to resolve hiatus, or (2) the reconstruction of PHEC ***ħ** (a voiceless pharyngeal fricative) with regular reflexes and some sporadic cases. The reconstruction analysis gains support from EC correspondences and reflex distribution patterns. The following data establishes the correspondence patterns, followed by the details of each interpretation:

	[520]		[163]		[519]	
PHEC	***ħuy ʔ-m**	‘be full’	***ħur-ʔ-**	‘young ox, bull’	***ħililll-**	‘smoke’
Burji	**huum-**	‘become full’	**(h)arʔeyi**	‘ox’	**hilili**	‘smoke’
PNHEC	***ħuy ʔ-m-**	‘be full’	***ħur-ʔ-**	‘young ox, bull’	***ħililll-**	‘smoke’
Had	**woʔm-**	‘be full’	**woʔla**	‘calf’	**wiriira, wuriira**	‘smoke’
Kam	**wiim-, yiim-**	‘be full’	**worjaamu**	‘young bull’	**wiliili-ta**	‘smoke’
Sid	**woʔma, hom-**	‘all, full’	**war-aamo**	‘ox, bull’	**wiliile**	‘smoke’
Ged	**woʔm-**	‘be full’	**-**	-	**wiliille**	‘smoke’

The reflexes show a consistent pattern: Burji changes ***ħ** to **h** initially, while the NHEC languages develop **w** before high vowels. The reflex **y** is attested dialectally in Kambaata [520] alongside **w**. Treis (2008, p. 39) found that, in contrast with the Duuraame dialect, Daambooyya replaces **w-** with **y-** before long front vowels.

Duuraame	Daambooyya	
wiimú	*yiimú*	‘to fill’
wiitú	*yiitú*	‘to grind’
weeʃʃú	*yeeʃʃú*	‘enset plant’ (sg)

Usually, the **h** reflex is optionally dropped before the vowel **a** in Burji, as in [163] *(h)arʔeyi* ‘young ox, bull’.[64] Sidaama *hom-* ~ *wom* ‘be(come) full’ is further evidence for positing a unique phoneme for such a series.

The pattern observed in [520], [163], and [519] is confirmed by additional attestations in the same environment. Before ***u**, Burji consistently shows **h** and NHEC shows **w**: PHEC [607] **ħuy ʔ-s-* ‘to fill’, [207] **ħun-*, **ħut-* ‘strength, power’, [208] **ħur ab-* ‘to pour’, and [1030] **ħuk’-* ‘be(come) thin’. Six examples establish this correspondence as secure.

Outside of initial position before the high vowels ***u** and ***i**, the reflexes show environment-dependent variation. The correspondence sets below illustrate these context-sensitive patterns, which require analysis of each vocalic environment separately.

> PHEC [467] **ħons-* ‘nine’ → Burji *wonfa* ‘nine’; PNHEC **ħons-* ‘nine’ → Had *honso* ‘nine’, Kam *honsu* ‘nine’, Sid *honse* ‘nine’.

> PHEC [698] **ħokol-* ‘to roll’ → Burji *kooliy-*, *kocikools-* ‘to roll’; PNHEC **ħok~kol-* ‘to roll’ → Had *onkoroll-*, *(h)onkoroll-* ‘to roll’, Kam *onkoloʔl-* ‘roll’, Sid *konkolati* ‘to roll’ (see Oro *konkolaad’d’a* ‘roll, roll down hill’).

Before other word-initial vowels (e.g., **o**), the only reasonable regularity is in the Hadiyya reflexes : [467] *honso* ‘nine’ and *(h)onkoroll-* ‘to roll’, suggesting NHEC **h** before **o** for ***ħ**. Otherwise, Burji is the one with the **w** reflex in Burji *wonfa* ‘nine’, an anomaly expected in a numeral. The only other example before **o** has even more sporadic reflexes: PHEC [698] **ħokol-* ‘to roll’ with assimilation to the flowing velar in Sidaama, **h ~ ø** variation in Hadiyya, loss in Kambaata, and apheresis (loss of initial syllable) in Burji.

The non-initial reflexes exhibit different series of correspondences in the following cognate sets:

> PHEC [439] **reħ-* ‘death’ → Burji *reya*; *reeya*, *reeha*, oya ‘death’; PNHEC **reħ-* ‘death’ → Had *leho* ‘death’, Kam *rehu-ta* ‘death’, Sid *reo*, *rewo*, *reo* ‘death’, Ged *reyo*, *reo* ‘death’.

[64] See the feminine set: Burji *hil-icco*, *hil-aanno*, Had *weeʔ-icco*, *waʔla*, Kam *w(aʔ)eccu-ta*, and Sid *waʔr-icco* ‘calf, young’ (Hudson, 1989, p. 35).

PHEC [184] **reħ ʔ* ‘die’ → Burji *re-*, *rey-* ‘die, be humilated’; PNHEC **reħ* ʔ ‘die’ → Had *leh-* ‘die’, Kam *reh-* ‘die, fall (of prices)’, Sid *re-*, *reʔ-* ‘die’, Ged *rey-*, *re(ʔ)-* ‘die’.

PHEC [329] **reħ-ʃ-* ‘corpse’ → Burji *reesi*, *reesi* ‘corpse’; PNHEC **reħ-ʃ-* ‘corpse’ → Had *leeʃa*; *leeʃʃa* ‘corpse’, Kam *re(e)ʃa* ‘corpse’, Sid *reeʃʃa*, *reeʃa* ‘corpse’, Ged *reenʃa*, *renʃa* ‘corpse’, Oro *reefa* ‘corpse’.

PNHEC [627] **zoħ ʔ* ‘to wander’ → Had *doʔ-* ‘roam, wander’, Kam *zoh-* ‘emigrate’, Sid *do-* ‘encircle, go round, wander’, Ged *zoor-* ‘roam, wander’.

PHEC [12] **maħ-* ‘what?’ → Burji *miyyaa*, *miya* ‘what?’; PNHEC **maħ* ‘what?’ → Had *ma*, *maha* ‘what?’, Kam *ma*, *maha*, *maʔa* ‘what?’, Sid *ma*, *maa*, *mai* ‘what?’, Ged *maa* ‘what?’.

PHEC [443] **uħw-* ‘to give’ → Burji *u(u)ww-* ‘give’; PNHEC **uħw-* ‘to give’ → Had *uww-* ‘give’, Sid *ʔui*, *aa*, *uww-* ‘give’, Ged *uww-* ‘give’.

The non-initial reflexes, as seen in [627], [329], [184], [439], and [12], show mostly Burji **ø** (or epenthetic **-y-**), Kambaata/Hadiyya **h**, and Sidaama/Gedeo **ø** (or epenthetic **-y-**/**-w-**). Such traces distinguish PHEC ***ħ** from PHEC ***h**, with the latter being completely lost in non-initial position. The alternation **y ~ ø** in non-initial reflexes is documented across multiple sources, though transcription practices vary. The following list [439] demonstrates that this variation may reflect linguistic reality in Burji, Sidaama, and Gedeo rather than inconsistent documentation:

Burji *reya* (Roba Dame & Wedekind, 2008, p. 69); *reeya, reeha* (Hudson, 1989, p. 47) ‘death’,

Sidaama *reo* (Yri & Pepper, 2019; Hudson, 1989, p. 47); *rewo* (Indrias Xa’miso et al., 2007, p. 489) ‘death’,

Gedeo *reyo* (Wedekind, 2008, p. 33); *reo* (Hudson, 1989, p. 47) ‘death’.

This cross-source consistency confirms the alternation as a synchronic phonological process in Gedeo (Eyob Kelemework, 2015, p. 36), suggesting incomplete loss of the pharyngeal reflex.

External comparative evidence strongly supports reconstruction of PHEC ***ħ**. Somali, within EC, shows **ħ** precisely where HEC languages have the variable reflexes **h**/**ø**/**y**. This correspondence pattern across EC indicates inheritance of ***ħ** from PEC, preserved in Somali, Afar, Saho, and Dullay, but variably reflected in HEC:

	PHEC	Burji	Sidaama	Somali	Gloss
[68]	***malaħ-**	**malaa**	**malaa**	**malaħ**	'pus'
[38]	***liħ-**	**lia, liya**	**lee, leye**	**liħ**	'six'
[20]	***sazeħ-**	**fadia, fadiya**	**sase**	**saddeħ, sadeħ**	'three'
[71]	***k'utaħ-**	**k'uwa, k'uwwa**	**uta**	**qodaħ**	'thorn'
[22]	***ħanʔur-**	**(h)anʔuroo**	**urco**	**ħunɖur**	'navel'

Additionally, PHEC ***ħ** as the reflex of PEC ***ħ** explains the persistence of identical final **a** across all the HEC in cognates like *mala* 'pus' from (PHEC (and PEC) [68] **malaħ-* 'pus' (§2.2.3.1, Terminal -a).[65] The systematic correlation between HEC reflex variation and Somali **ħ** undermines an epenthesis-only account, which would not predict cross-linguistic correspondence.

However, two EC parallels show correspondences other than **ħ** for the regular initial series (1): PHEC ***ħ** → Burji **h**, NHEC **w** before **u**:

> PHEC [163] **ħur-ʔ-* 'young ox, bull': Saho *awur* 'bull, animal for mounting', Afar *awur* 'bull, ox', Som *awr* 'male camel', and Das *ʔauric* 'non-castrated camel', Ren *óor* 'bull (of camel or cattle)', Maay *ọọr* 'male camel', Boni *óor* 'male elephant', Arb *ʔáar* 'bull', Elm *aar* 'male crocodile', Bay *aar* 'bull, ox'.
>
> PHEC [208] **ħur ab-* 'to pour': Bay *woraab-* 'drink water, be watered'; Ren *warab-* 'draw water', *wor* 'waterhole, borehole', Som *waraab-i-* 'to water', *wár(-tii)* pool, pond', Maay *wạr* 'swamp', Arb *war(a)b-* 'draw and fetch water', Das *wár* 'river (the Omo)', Kon *oraap-* 'fetch, dip, pour (water only?)', Yaaku *irp-a-* 'draw water'.

[65] The reflexes of the PEC voiced pharyngeal ***ʕ** do not show unique behavior when compared to the reflexes of PHEC ***ʔ**, e.g., PEC [30] **kaʕ-* 'to get up, arise' → PHEC **kaʔ-*, **keʔ-* 'to get up, arise', so they have likely merged. PHEC [519] **ħilill-* 'smoke' has parallels in Afar *ʕir*, *ʕer* 'smoke' and *ʕeèri* 'steam' with an unexpected voiced pharyngeal **ʕ** other than **ħ**, so if they are cognates, an intermediate ***ħ** or irregular Afar **ʕ** must be posited to explain the PHEC correspondence.

Many of the reflexes are comparable to those of NHEC with **w**, and others show mostly **ø** or a rounded vowel, proving the establishment of the **u** environment. Since they are only two examples, their implications with respect to the PEC reconstructions will need to be considered, for instance: (a) they might be reflexes of a consonant other than ***ħ**, e.g., a voiced glottal fricative ***ɦ**, or (b) they might be examples of ***ħ** change conditioned by the following **u** as occurs in NHEC. For now, the latter hypothesis is satisfactory for addressing the HEC reflexes, given the limited parallels with EC. However, a wider EC reconstruction might require considering the additional phoneme hypothesis if supported by further evidence. Here is one comparative set supporting the reconstruction of ***ħ** to address such a challenge:

> PHEC [139] **wozan-* 'heart' → Burji *wodana* 'heart'; PNHEC **wozan-* 'heart' → Had *wodano*, *wadno* 'heart', Kam *wozana*; *wazana* 'heart, mind', Sid *wodana* 'mind (heart), human heart', Ged *onna* 'heart, courage', Oro *onn-ee* 'heart, courage', Kon *otan-ta* 'center', Saho *wazan-a* 'heart', Bay *wadan* 'heart', Ren *wˈéyna* 'heart', Som *wadne*, *wadane* 'heart', Maay *wenni* 'heart', Boni *wˈéene* 'heart', Das *ʰuoðin-ni* 'heart'.

The reconstruction of PHEC **wozan-* 'heart' is without issue as long as the current hypothesis of Oromo influence on Gedeo *onna* 'heart, courage' is accepted as an explanation for the unexpected loss of initial **w**. Even in that case, however, the Oromo cognate *onn-ee* 'heart, courage' and Konso *otan-ta* 'center' would still require explanation. The reconstruction of an old EC root **ɦuzn-*, **ɦuzan-* 'heart' might help address those issues and is well supported by the superscripted initial in the Dhasanach cognate *ʰuoðin-ni* 'heart'.[66] Alternatively, the analysis could fall back on the parsimonious hypothesis of PEC ***ħu** yielding **wV** and **øV** in EC with an irregular **w** in Burji.

The lack of a reflex of initial ***h** before the back high vowel ***u** in HEC is a possible counterargument against positing a new phoneme. Still, it is justifiable to argue that (a) the reflexes of the phoneme ***ħ** in word-initial position are not restricted to the environment before ***u**, as shown in [519] and [698], and (b) PHEC ***hu** might have merged with ***ħu** and resulted in Burji **hu**, NHEC **wV**.

[66] The assimilation of the initial fricative ***ħ** to the second fricative **z** would also accommodate the Arbore cognate *zazzá* 'heart' as more plausible than the assimilation of a glide **w**.

This is not the only merger – an earlier ***ħa*** merging into ***ha*** would account for examples like the following:

> PHEC [426] **ħaraar y-s-, *araar y-s-* ‘to mediate, reconcile’ → Burji *(h)araar-s-* ‘mediate, reconcile’; PNHEC **ħaraar y-s-, *araar y-s-* ‘to mediate, reconcile’ → Had *araar-*, *araass-* ‘to reconcile, relieve (from danger)’, Kam *araar-s-* ‘mediate, intervene’, Sid *araar-s-* ‘mediate, reconcile’, Ged *araar-s-* ‘mediate, reconcile’ (see Oro *araarsa* ‘mediate, reconcile’).

The reconstruction of initial ***ħ*** in [426] is tentatively based on the free variation of initial **h ~ ø** in Burji *(h)araar-s-* ‘mediate, reconcile’. This makes the alternative reconstruction **araar-s* ‘to mediate, reconcile’ also reasonable by arguing for an analysis of areal epenthesis to explain the variation in Burji.[67] The lack of ***ħa*** and the frequent attestation of the PHEC sequence ***ha*** (§2.1.5.2) also support the assumption of a merger of ***ħa*** with ***ha*** at an early stage, making [426] an exception.

PHEC	Merged		Burji	Had	Kam	Sid	Ged	Environment
	***h**	→	**h/ø**	**h/ø**	**h/ø**	**h/ø**	**h/ø**	**#_a**
***ħ** and ***h**	→							
	***ħ**	→	**h**	**w**	**w**	**w**	**w**	**#_u**

The **h ~ ø** variation in the reflexes of the merger before **a** may be a later areal phenomenon and includes the reflexes of ***h**.

As mentioned above, two competing analyses present themselves. The first treats Burji **h** and PNHEC **w/y** as epenthetic segments inserted before certain vowel-initial stems. The epenthesis account would derive **w**/**y** from hiatus resolution. This analysis has precedents in synchronic Kambaata/Sidaama morphophonology (§3.4.1.2). However, three factors challenge mere epenthesis:

(a) Distribution asymmetry – epenthesis typically resolves specific morpheme boundaries, but the **h**/**w** reflexes appear in the roots.

[67] The variation between **h** and **ø** in initial position has been reported for Burji (Tesfaye Baye, 2015, p. 78) and Kambaata (Treis, 2008, p. 27). Treis (2008, pp. 27, 32) states that initial **h ~ ʔ** variation, e.g., Kambaata *hank’afu ~ ʔank’afu* ‘to embrace’ is widespread in East Cushitic and Ethiosemitic languages. However, she stresses that there is no general lack of distinction between **h** and **ʔ** word-initially. Such areal initial epenthetic **h-** is more prevalent in Gedeo and is even attested in established vowel-initial stems, e.g., pronouns.

(b) External cognates such as Somali **ħ** in cognate items suggest an inherited segment rather than independent innovation.
(c) Divergent Burji/NHEC reflexes weaken the analysis as epenthesis, since it would predict parallel outcomes; the regular **h**/**w** split suggests proto-segment conditioning.

The epenthesis mechanism may explain some examples (particularly those lacking external cognates), but cannot account for the full distribution.

The second analysis reconstructs a proto-phoneme ***ħ** with the reflexes presented above. Several methodological considerations favor this analysis.[68]

First, the principle of economy favors the reconstruction of a single proto-phoneme over the positing of two independent epenthesis processes (one in Burji, another in PNHEC) that happen to align perfectly across the attested cognate sets. Second, the comparative method's uniqueness principle holds that a unique correspondence pattern should be analyzed as reflecting a unique proto-phoneme unless there is compelling evidence to the contrary. Third, the EC parallels show a mostly regular correspondence with **ħ** in those languages that retain the pharyngeal. Fourth, the PNHEC **w** ~ **y** variation (as in Kambaata [519]) is best understood as phonetically natural assimilation of ***ħ** to the following vowel, and alternations in Sidaama *hom-* ~ *wom-* 'be full' suggest an underlying segment rather than epenthetic instability. This is what we expect from an inherited segment undergoing conditioned sound change. Fifth, the epenthesis hypothesis faces the empirical challenge that word-initial epenthesis contradicts the otherwise inter-morphemic position of all other HEC epenthetic processes.

Additional support comes from the vowel quality variation observed across the PNHEC and EC languages in these roots. The alternation between Hadiyya, Sidaama, Gedeo *woʔm-* and Kambaata *wiim-* 'be(come) full' in [520], and more generally the varied vowel qualities in the PNHEC reflexes, may reflect coloring effects from an earlier voiced glottal fricative that lowered or backed adjacent vowels before and after merging with the glides. Furthermore, such vowel variation is common in other EC languages (e.g., WOT) when they occur adjacent to the pharyngeals.

[68] See exceptions such as Somali [256] *ħan* 'slander, gossip' and [229] *ħanʃaʃaq* 'whisper', which are reconstructable with regular **h** reflexes from PHEC **ham ʔ* 'to slander' and **haʃ~haʃ-* 'to whisper', respectively.

The following list evaluates the reflexes of PHEC ***ħ** attested in different environments.

	Position	Burji	Had	Kam	Sid	Ged	Environment	Evaluation
1.	Initial	**h**	**w**	**w**	**w**	**w**	**/__*u, (*i)**	secure
2.	Initial	**w/ø**	**h**	**h**	**h**	**h**	**/__*o**	tentative
3.	Initial	**h ~ ø**	**ø**	**ø**	**ø**	**ø**	**/__*a**	tentative
4.	Non-initial	**y/ø**	**h**	**h**	**ø**	**y/ø**	**/V__**	reasonable

The observed tendencies include: (1) regular NHEC **w** and Burji **h** before ***u** (and maybe ***i**), (2) sporadic NHEC **h** and Burji **ø**/**w** before ***o**, (3) tentative NHEC **ø** and Burji free variants **h ~ ø** word-initially before ***a**, following the **ħ** and **h** merger,[69] and (4) sporadic NHEC **h**, **ø** and Burji **ø** in non-initial/post-vocalic position (Burji and Gedeo **y** might be reflexes, but the hiatus that a zero reflex creates is also resolvable by epenthetic **-y-/-w-**, see §§3.4.1.2, 3.3.3).

Series (1) is evaluated as 'secure' based on six supporting examples with consistent reflexes across all five HEC languages. Series (2) and (3) are 'tentative' due to limited attestation (two and one examples respectively) and irregular reflexes in individual languages (Burji [467] *wonfa* 'nine' showing **w** rather than the expected **h** before **o**). Series (4) achieves 'reasonable' status despite variation (Burji/Gedeo **y**/**ø** vs. Kambaata/Hadiyya **h**) because the contrast with complete loss of ***h** in non-initial position distinguishes ***ħ** from ***h** diachronically.

In sum, PHEC ***ħ** is arguably reconstructable on the basis of a regular correspondence set showing Burji **h**/**ø**: PNHEC **h**/**w**/**ø**. While synchronically unattested, its reconstruction is motivated by principles of economy, EC correspondences, and the comparative method. This analysis brings the reconstruction of HEC into closer alignment with standard comparative methodology, eliminating the need to posit word-initial epenthesis processes that are otherwise unattested in the language family. The reconstruction raises the PHEC consonant inventory from 20 to 21 segments.

[69] Before the low vowel **a**, the reflexes are in free variation **h ~ ø** in Burji (supported by the dictionary listing in Getahun Addo (2017)) and mostly **ø** in NHEC with sporadic **h** in some lexemes.

2.1.5.4 Summary

The velar, pharyngeal, and glottal consonants bring together two themes that recur throughout this chapter: the asymmetry between initial and non-initial positions, and the fact that several "phonological" outcomes are only intelligible once morphophonological context is taken into account.

The reconstruction reveals a contrast between stable initial retention and significant non-initial weakening. The velar stops ***g** and ***k** are securely reconstructed, though they exhibit divergent developments: ***g** is susceptible to palatalization in Burji, while ***k** is frequently lost or lenited in non-initial position unless protected by gemination or clustering. Non-pulmonic ***k'** presents a more complex picture. It is reconstructable for PNHEC but shows signs of marginal status or reintroduction through contact in the modern languages. Among the glottal consonants, ***h** is stable initially but prone to loss elsewhere, and the comparative evidence supports reconstructing ***ʔ** as part of the stem in a subset of items alongside its multiple grammatical functions (genitive/reflexive/nominal and epenthetic uses). The arguable reconstruction of the pharyngeal ***ħ** is supported by distinctive internal and EC correspondences. Together, these patterns signify the role of both internal lenition processes and external contact influence in shaping the HEC phonological inventory.

Finally, this section has already foreshadowed two subgroup-relevant developments that will be relevant again in §2.3: the relative instability of certain onsets in the northern branch of HEC, and the way glottal-related alternations feed into later restructuring.

2.2 HEC Vowels

The HEC languages have five vowel qualities, each contrasting between single and double quantity.

Table 2: *Highland East Cushitic Vowels Inventory*

	Front		Back
High	**i, ii**		**u, uu**
Mid	**e, ee**		**o, oo**
Low		**a, aa**	

Following a moraic analysis, I refer to **single V** and **double VV** (for identical-vowel sequences) rather than "short" vs. "long" vowels. All HEC languages

have five single and double vowels (Hudson, 1976, p. 247). Section 2.2.1 treats regular reflexes; Section 2.2.2 addresses deviations in quantity and quality.

2.2.1 Regular Stem Vowels

The sections below present the regular reflexes of stem single vowels across the five HEC languages. They are all securely reconstructed for PHEC and PNHEC. Section 2.2.2 treats deviations in quantity or quality separately, since many of them reflect later changes rather than inherited vowel contrasts.

2.2.1.1 Single Vowels

There are five reconstructable HEC single vowels: ***i**, ***e**, ***a**, ***o**, ***u**. Some reflexes of the single vowels exhibit irregularities in the corresponding vowel quantities depending on the language.

PHEC *i

	[1]		[54]		[702]	
PHEC	***ill-**	‘eye’	***min-**	‘house’	***hig-**	‘to return, pass’
Burji	**illa**	‘eye’	**mina**	‘house’	**hig-**	‘exchange’
PNHEC	***ill-**	‘eye’	***min-**	‘house’	***hig-**	‘to return, pass’
Had	**ille**	‘eye’	**mine**	‘house’	**hig-**	‘pass’
Kam	**illi-ta**	‘eye’	**mini**	‘house’	**hig-**	‘pass’
Sid	**ʔille**	‘eye(s)’	**mine**	‘house’	**hig-**	‘return’
Ged	**ille**	‘eye(s)’	**mine**	‘house’	**hig-**	‘return, repeat’

Single ***i** is also found in sets such as PHEC [151]**is-e* → PNHEC **ise* ‘her’, PHEC [7] **is-u* → PNHEC **is-u* ‘he’, PHEC [46] **ik’k’-*, **ikk-* → PNHEC **ink’-*, **ink-* ‘tooth’, PHEC [237] **ʃirb* → PNHEC **sirb-* ‘to dance and sing’, PHEC [287] **t’ib- (ʔ)* → PNHEC **t’ib-* ‘to make trouble’, PHEC [392] **lik’-* → PNHEC **lik’-* ‘to lend’, and PHEC [637] **sim-* → PNHEC **sim-* ‘ensete tree 1st stage’. However, it is also attested in some series of irregular correspondences discussed in §2.2.2.

PHEC *e

	[184]		[407]		[223]	
PHEC	***reħ-**	‘die’	***ebel-, *ebal**	‘so-and-so’	***eg-**	‘to wait’
Burji	**rey-**	‘die’	**ebeloo**	‘so-and-so’	**eeg-**	‘to wait’
PNHEC	***reħ ʔ**	‘die’	***ebel-, *ebal-**	‘so-and-so’	***eger**	‘to wait’
Had	**leh-**	‘die’	**ebaro**	‘so-and-so’	**eger-, egar-**	‘wait (for)’
Kam	**reh-**	‘die’	**ebalo**	‘so-and-so’	**eger-, agar-**	‘wait (for)’
Sid	**re-, reʔ-**	‘die’	**ʔebelo**	‘so-and-so’	**ʔagar-**	‘wait, watch’
Ged	**rey-, re(ʔ)-**	‘die’	**ebelo**	‘so-and-so’	**egert-, eged’-**	‘wait!, wait (for)’

The disyllabic examples above show that **e** and **a** can sporadically assimilate, such as [223] in all NHEC and [407] in Gedeo and Sidaama, or dissimilate, as in Hadiyya and Kambaata [407] (← **ebel-*) and [223] (← **eger*).[70] It is more common for disyllabic stems to have identical vowels, which supports the dissimilation proposal (consistent with the East Cushitic echo-vowel epenthesis (Sasse, 1979)).

Additional examples include PHEC [114] **debeʔ-* → PNHEC **debeʔ-* ‘to belch’, PNHEC [75] **t’eʔm-* ‘to taste good’, PHEC [144] **meʔ-* → PNHEC **meʔ-* ‘how many? how much?’, PHEC [288] **k’el-* → PNHEC **k’el-* ‘snail’, PHEC [514] **ebb-* → PNHEC **ebb-* ‘to bring’. Overall, single ***e*** is found in limited comparanda in HEC.

Further examples show some irregularity in the quantity of the corresponding vowels due to later morphophonological processes (§2.2.2.2). For instance, PHEC verb [184] **reħ ʔ* ‘die’ above is derived from the same stem as the noun PHEC [329] **reħ-ʃ-*, the PNHEC **reħ-ʃ-* ‘corpse’; however, other than Kambaata *reʃa* ‘corpse’, all HEC noun cognates have reflexes with double **ee**: Burji *reesi* ‘corpse’, Had *leeʃa* ‘corpse’, Sid *reeʃʃa* ‘corpse’, Ged *reenʃa* ‘corpse’ (see Oro *reefa* ‘corpse’). The vowel doubling is attributed to the loss of the glottal **h** in that noun.

[70] PNHEC [223] **eger* correspondences include the unexplainable **-er** which has no parallel in the rest of EC (see PEC **ʔeeg-* ‘watch, look out’ (Sasse, 1982, p. 67)).

For examples of an alternation between **e** and **a** before **ʔ**, see PHEC [269] **t'eʔ-* 'to diminish', **t'aʔ-* 'to extinguish' (same set with alternating **e ~ a**) and PHEC [30] **kaʔ-* → PNHEC **kaʔ-*, **keʔ-* 'to get up, arise' (§2.2.2.1).

PHEC *a

	[137]		**[2]**		**[189]**	
PHEC	***lal-**	'cattle'	***an-i**	'I'	***k'araar-**	'be bitter'
Burji	**lál-i**	'cattle'	**ʔáṇ̩**	'I'	**k'araar-**	'be bitter'
PNHEC	***lal-**	'cattle'	***ani**	'I'	***k'araar-**	'be bitter'
Had	**laro**	'cattle'	**ʔáṇ̩**	'I'	**k'araar-**	'be bitter'
Kam	**lalu**	'cattle'	**án**	'I'	**k'araar-**	'be bitter'
Sid	**laló**	'cattle'	**ʔani**	'I'	**k'araar-**	'be bitter'
Ged	**lalo**	'cattle'	**ani**	'I'	**k'araar-**	'be bitter'

Further examples of single PHEC ***a** include: PHEC [80] **gal-* → PNHEC **gal-* 'to enter', PHEC [9] **arrab-* → PNHEC **arrab-* 'tongue', PHEC [14]**az-* → PNHEC **az-* 'milk', PHEC [36] **gam-* 'to bite' → PNHEC **gam-ʔ-* → **gaʔm-* 'to bite', PHEC [67] **ang-* → PNHEC **ang-* 'hand, arm', PHEC [68] **malaħ-* → PNHEC **malaħ-* 'pus', PHEC [72] **lam-* → PNHEC **lam-* 'two', PHEC [110] **bar-* → PNHEC **barr-* 'day', PHEC [137] **lal-* → PNHEC **lal-* 'cattle', and PHEC [449] **k'al-* → PNHEC **k'al-* 'give birth, beget (of animals)'.

Gedeo additionally has a reflex **e** of PHEC ***a**. Such a change of ***a** to **e** is observed in suffixes such as **-em** (pass) from PHEC ***-am** (§4.2.1) and **-ed'** (mid) from Oromo **-ad'** (§4.2.3). An example of the passive suffix is Gedeo [545] *dun-em-* 'leak (of e.g. water)'. Wedekind (1980, pp. 139–149) relates the correspondence between **a** and **e** to accent placement in Gedeo (§2.4), where the final syllable of the stem is stressed, while most suffixes and some nominal and verbal forms require penultimate stress.

	PNHEC		**Gedeo**	
[545]	***dun-am-**	'leak (of water)'	**dun-em-**	'leak (of water)'
[90]	***malab-**	'honey'	**malebo**	'honey'
[444]	***bagaz-**	'spear'	**bagedo, bagado**	'spear'
[707]	***magan-**	'sky, God'	**mageno, magano**	'God, sky'
[59]	***agan-**	'month, moon'	**agen-jo**	'month, moon'
[10]	***hat't'iʃ y**	'to sneeze'	**het't'if-at-**	'sneeze'

In the above examples, Gedeo **e** is found in the penultimate syllable, except in example [10]. The variation between **e** and **a** in [444] and [707] is a difference in the source's transcription: [444] *bagedo* (Wedekind, 2008, p. 74),

bagado (Hudson, 1989, p. 139), [707] *mageno* (Wedekind, 2008, p. 35), *mageno*, *magano* (Hudson, 1989, p. 71).

As Sasse (1982, p. 16) notes, the contrast between **a** and **e** is neutralized in closed syllables before **y** in Burji. That results in a difference in transcription depending on the source, e.g., sometimes Sasse (1982) has **a** where Straube's data has **e**.

PHEC *o

	[478]		[139]		[580]	
PHEC	***kott-**	'hoof, foot'	***wozan-**	'heart'	***boc'-**	'to carve, split'
Burji	**kottee**	'hoof'	**wodana**	'heart'	**boc'-**	'carve'
PNHEC	***kott-**	'hoof, foot'	***wozan-**	'heart'	***boc'-**	'to carve, split'
Had	**konteʔe**	'hoof, foot'	**wodano**	'heart'	**boc'-**	'carve, to split'
Kam	**kotti-ta**	'foot step'	**wozana**	'heart'	**boc'-**	'snap (to break)'
Sid	**kotte**	'hoof'	**wodana**	'heart'	**boc'-**	'carve, split'
Ged	**kotte**	'hoof'	**(onna)**	'heart'	**boc'-**	'carve, split '

Further examples of single PHEC ***o*** include: PHEC [126] **gog-* → PNHEC **gog-* 'skin, hide', PHEC [87] **k'ot-* → PNHEC **k'ot-* 'to dig', PHEC [95] *ʃomb-* 'lung', and PHEC [286] **c'ob-* → PNHEC **c'ob-ʔ-* → **c'op'p'y-* 'to drip, to leak'.

PHEC *u

	[48]		[300]		[108]	
PHEC	***tum-**	'to pound'	***k'ub-**	'finger(-ring)'	***sunk'**	'to kiss'
Burji	**tum-**	'to pound'	**k'ubee**	'finger-ring'	**sunk'-**	'kiss'
PNHEC	***tum-**	'to pound'	***k'ub-**	'finger(-ring)'	***sunk'**	'to kiss'
Had	**tum-**	'pound (metal)'	**k'ubaʔa**	'finger-ring'	**sunk'-**	'kiss'
Kam	**tum-**	'pound (metal)'	**k'ubbata**	'finger-ring'	**sunk'-**	'kiss'
Sid	**tum-**	'hit, strike'	**k'ubb-e**	'ring'	**sunk'-**	'kiss'
Ged	**tum-**	'pound (metal)'	**k'ube**	'ring'	**sunk'-**	'kiss'

The meanings of [300] **k'ub-* include 'finger(-ring)' and 'ring'.

Further examples of single PHEC ***u*** include: PHEC [40] **tuf-* → PNHEC **tuf-* 'to spit', PHEC [672] **dul-* → PNHEC **dul-* 'to skin/slaughter', PHEC [180] **utub-* → PNHEC **utub-* 'pole, center-pole of house', PHEC [408] **mut-* → PNHEC **mut-* 'awl for basketwork', PHEC [268] **mur-* → PNHEC **mur-* 'to cut', PHEC [224] **t'ur-* '→ PNHEC **t'ur-* 'dirt of body, clothing', PHEC [391] **kum-* → PNHEC **kum-* 'thousand', and PHEC [381] **k'ul-*, **k'ul~k'ul~l-ʔ-* → PNHEC **ul-*, **ul~ul~l-ʔ* 'leech'.

2.2.1.2 Double Vowels

PHEC shows evidence for double-vowel counterparts to every single vowel (***ii**, ***ee**, ***aa**, ***oo**, ***uu**). They are secure reconstructions, though these are less frequent in the current database than single vowels.

PHEC *ii

	[45]		[400]		[250]	
PHEC	***fiik' ʔ**	'to whistle'	***giir-**	'fire'	***t'iib-**	'to push'
Burji	**fiisk'-**	'whistle'	**jiira**	'fire'	**d'iib-**	'to push'
PNHEC	***fiik' ʔ**	'to whistle'	***giir-**	'fire'	***t'iib-**	'to push'
Had	**fiink'ees-**	'whistle'	**giira**	'fire'	**t'iib-**	'to push'
Kam	**fiik'-**	'whistle'	**giira-ta**	'fire'	**t'iib-**	'press'
Sid	**fiik'a**	'whistle'	**giira**	'fire'	**t'iibi**	'push'
Ged	**fiiss-**	'whistle'	**giira**	'fire'	**d'iib-ema**	'be pushed'

Additional examples of double PHEC ***ii*** include: PHEC [390] *hiit'-* → PNHEC *hiit'-* 'to stretch', PHEC [348] **c'iiʔ-* 'bird' → PNHEC **c'iiʔ-i-ʃʃ-*, PHEC [493] **hiiʔ-* → PNHEC **hiiʔ-* 'take!', PHEC [273] **biin-* 'mosquito, fly (sp.)' → PNHEC **biin-* 'mosquito, fly (sp.)', PHEC [652] **kiil-* → PNHEC **kiil-* 'to do magic', and PNHEC [115] **k'iid-* 'become cold'.

One set remains difficult: [583] is reconstructed as PHEC/PNHEC **maaz ʔ* alongside**miiz ʔ* 'to hurt'. The distribution is split, with ***ii*** in Burji *miid'-* 'to harm, hurt', Sidaama *miic'i* 'hurt', and Gedeo *miid'a* 'hurt', but ***aa*** in Hadiyya *madaʔ-* 'to wound somebody' and Kambaata *maaz-*, *maz-ees-* 'be wounded, wound somebody'. This could reflect (a) dialect mixture at the PNHEC stage, (b) analogical remodeling in one sub-branch, or (c) contact influence, as suggested by the Oromo parallel *miid'a* 'hurt'.

PHEC *ee

	[462]		[613]		[136]	
PHEC	***meet-**	'scraping board'	***t'een-**	'rain'	***eel-**	'pond, pool'
Burji	**meeta**	'scraping board'	**c'eena**	'rain'	**eela**	'water-hole'
PNHEC	***meet-**	'scraping board'	***t'een-**	'rain'	***eel-**	'pond, pool'
Had	**meeta**	'scraping board'	**t'eena**	'rain'	**leera**	'pond, pool'
Kam	**meeta**	'scraping board'	**t'eena**	'rain'	**eela**	'pond, pool'
Sid	**meeta**	'scraping board'	**t'eena**	'rain'	**ʔeela**	'low place, pond'
Ged	**meeta**	'scraping board'	**t'eena**	'rain'	**eela**	'puddle, water'

Further examples of ***ee** are mostly reconstructable for NHEC: PHEC [326] **geeg-eess-* → PNHEC **geeg-eess-* 'accompany' (see Oromo *geggeessa* 'accompany'), PHEC [79] **eey-* → PNHEC **ee*, **yee* 'yes', PNHEC [42] **k'eer-* 'long, tall', PNHEC [406] **wees-* 'ensete tree (Ensete ventricosum)', PNHEC [460] **beet-* 'son, child', PNHEC [741] **beet-*, **beet-beet-* 'grandchild', and PNHEC [437] **seer-* 'law'.

Similar to single ***e**, there are examples of double ***ee** with some irregularity in quantity and sometimes in the quality of the corresponding vowels, e.g., PHEC [168] **hee?-* 'to exist, to live' is transcribed with double vowels in Hadiyya *hee?-* 'live, exist', Sidaama *heed'-* 'exist, be, be in place, be alive', and Burji *heettaa* 'here it is!', but a single vowel in Kambaata *he?-* 'live' and Gedeo *hed'-* 'exist, be there'. The following examples suggest a change from PHEC ***ee** to Burji **ay**, but there is no regular substantial evidence to establish such a rule:

	Burji		**PNHEC**	
[440]	**wayne, woynee**	'Colobus monkey'	***wee?n-**	'Colobus monkey'
[634]	**laymi**	'bamboo'	***leem-**	'bamboo'
[128]	**d'aybood'-**	'to become thirsty'	***t'eeb ?**	'to be thirsty'

PHEC *aa

	[189]		**[492]**		**[209]**	
PHEC	***k'araar-**	'be bitter'	***haar-**	'new'	***laaf-**	'soft, weak'
Burji	**k'araar-**	'be bitter'	**haareyaa**	'new'	**laafa**	'soft, weak'
PNHEC	***k'araar-**	'be bitter'	***haar-**	'new'	***laaf-**	'soft, weak'
Had	**k'araar-**	'be bitter'	**haareécco**	'new'	**laafa**	'weary, infirm'
Kam	**k'araar-**	'be bitter'	**haaroo**	'new'	**laafa**	'soft, weak'
Sid	**k'araar-**	'be bitter'	**haaro**	'new'	**laafa**	'weak'
Ged	**k'araar-**	'be bitter'	**haaro**	'new'	**laafa**	'soft, weak'

Similar to single ***a**, double ***aa** is frequently attested in the available database. Here are additional examples: PHEC [635] **maal-* → PNHEC **maal-* 'meat', PHEC [252] **hasaaw-* → PNHEC **hasaaw-* 'speak, talk', PHEC [258] **maak-*, **ma~mmaak-* → PNHEC **maak-* 'say in a proverb, tell a story', PHEC [284] **ayyaan-* → PNHEC **ayyaan-* 'festival, holiday' (also found in Ethiosemitic languages), PHEC [354] **yaa?-* → PNHEC **yaa?-* 'to flow, assemble', PHEC [450] **ʃaal-* → PNHEC **saal-* 'be shy, be ashamed', PHEC [548] **arraab-* → PNHEC **arraab-* 'to lick', PHEC [703] **kaakk y-s-* → PNHEC **kaakk y-s-* 'to cackle (hen)', PHEC [3] **aabb-* → PNHEC **aabb-* 'father', PHEC [150] **baall-* → PNHEC **baall-* 'feather', PHEC [426] **ħaraar y-s-, *araar y-s-* → PNHEC **ħaraar y-s-, *araar y-s-* 'to mediate, reconcile', and PHEC [282] **aar* 'be angry'.

Interestingly, it is mostly attested adjacent to the liquids **l** and **r**, and to a lesser extent, the other sonorants, but is least common between obstruents.

PHEC *oo

	[330]		[8]		[675]	
PHEC	***sooz-**	‘dawn’	***ʃool-**	‘four’	***k’oom-**	‘calabash’
Burji	**sood-**	‘dawn’	**foola**	‘four’	**k’oomoo**	‘calabash’
PNHEC	***sooz-**	‘dawn’	***ʃool-**	‘four’	***k’oom-**	‘calabash’
Had	**sood-**	‘dawn’	**sooro**	‘four, lot’	**k’ooma**	‘calabash, gourd’
Kam	**sooz-**	‘dawn’	**ʃoolu**	‘four’	**k’oomu**	‘calabash’
Sid	**sood-**	‘dawn’	**ʃoole**	‘four’	**k’oon-co**	‘calabash, skull’
Ged	**sood-**	‘dawn’	**ʃoole**	‘four’	**k’oon-jo**	‘gourd, vessel’

Additional examples of double ***oo*** include: PHEC [31] **oo y*, **oo ʔ* ‘to cry’ → PNHEC **oo y-*, **oo ʔ* ‘to cry’,[71] PNHEC [217] **fool-* ‘breath, odor, smell’, PHEC [610] **roob-* → PNHEC **roob-* ‘hippopotamus’, PHEC [603] **boos-* → PNHEC **boos-* ‘fireplace’, PHEC [623] **oot-* → PNHEC **oot-* ‘fence’, PNHEC [347] **oos-* ‘boy’.

Gedeo has a few words transcribed with a single **o** that correspond to double **oo** in the rest of NHEC. One example is the regular **oo** in [675] *k’oon-jo* ‘gourd, vessel, calabash’ above, which has a variant transcribed with a single **o** as seen below, in addition to the other two examples.

	PNHEC	**Gloss**	**Gedeo**	**Gloss**
[675]	***k’oom-**	‘calabash’	**k’omma**	‘gourd, vessel, calabash’
[222]	***t’ook-**	‘to burst’	**d’oʔ-**	‘burst, let burst; burst’
[327]	***koor-**	‘saddle’	**koricca**	‘saddle’

Burji and wider EC parallels point to a single **o** for [222] Burji *t’okk-* ‘to leak, drop, burst’, Oromo *d’oʔa* ‘burst, explode’, Afar *ɗokʕ-* ‘bruise, dent’, and in [327] Burji *koraa* ‘saddle’, Konso *kor-aa* ‘saddle’, Saho, and Afar *kor* ‘saddle’. These two sets are therefore reconstructed with a single vowel at PHEC level, i.e., [222] **t’ok-* ‘to burst’ and [327] **kor-* ‘saddle’. The short vowel in Gedeo need not, however, be inherited at PNHEC level. Given the recurrent Oromo influence on Gedeo, as argued elsewhere, a secondary short vowel is especially plausible in [222], which has a close Oromo parallel. I therefore treat the Gedeo corresponding vowel here as a local, likely contact-influenced reflex rather than as evidence for a regular PNHEC shortening rule. The

[71] Compound from PEC **ʔoo y-* ‘say oo’ (Sasse, 1982, p. 156).

PNHEC long vowels are retained on the basis of the non-Gedeo NHEC evidence.

PHEC *uu

	[351]		[438]		[325]	
PHEC	***t'uut'-**	'to suck'	***zuug-**	'scrape, tan'	***duud-**	'mute (person)'
Burji	**t'uunt'-**	'absorb'	**duug-**	'scrape, tan'	**duuda**	'dumb'
PNHEC	***t'uut'-**	'to suck'	***zuug-**	'scrape, tan'	***duud-**	'mute (person)'
Had	**t'uut'-**	'to suck'	**duug-**	'scrape, tan'	**duuda**	'mute'
Kam	**t'uut'-**	'to suck'	**zuug-**	'scrape, tan'	**duuda**	'mute'
Sid	**t'uut'-**	'to suck'	**duug-**	'scrape, tan'	**duuda**	'mute'
Ged	**t'uut'-**	'to suck'	**duug-**	'scrape, tan'	**duuda**	'blunt, dull'

Double ***uu** is less frequently attested than the other vowels. Additional examples include PHEC [130] **tuul-* → PNHEC **tuul-, *tuur-m-* 'to pile up', PNHEC [49] **k'uub-*, **k'uub-ʔ-* 'egg', PNHEC [411] **buur-* 'butter', and PNHEC [742] **buur-*, **buur-ʔ-* 'to anoint/smear, anoint oneself'.

Two examples show differences in the quantity of corresponding vowels:

> PHEC [66] **anuun-* 'breast' → Burji *ununa* 'breast'; PNHEC **anuun-* 'breast' → Had *anuuna* 'breast', Kam *anuuna* / *unuuna* 'breast', Sid *unuuná* 'breast', Ged *unuuna* 'breast'.
>
> PHEC [186] **c'ub-* 'dip in, immerse' → Burji *c'uub-* 'immerse, to dip, soak'; PNHEC **c'ub-*, **c'ub ʔ*, **c'ub y-s* 'to rinse, immerse' → Had *c'ufc'ufaaʔ-* 'to wash, to immerse', Kam *c'uf aʔ* 'rinse, immerse', Sid *c'uuʔi* 'baptise', Ged *c'uup'a* 'dip (in)'.

In [66], the double vowels are continued in PNHEC, but are transcribed with a single **u** in Burji. The initial vocalic alternation of **a** with **u** is an irregular assimilation, as the free variation of the two Kambaata variants suggests. The second example [186] is reconstructed with a single ***u** vowel, but it has a double **uu** in Sidaama, Gedeo, and Burji. These double vowels could be due to contact with Oromo (see *c'uup'a*, *c'up'-*, *c'ub-* 'dip, baptize') or internal alternation processes.

2.2.2 Variation in Stem Vowels

Several comparative sets show irregular correspondences in vowel quality and quantity. Where possible, I group these sets by conditioning environment, since at least some of the variation plausibly reflects later developments rather

than inherited contrasts. The likely sources of irregularity include adjacent consonants (especially glottals and labials), accent placement, morphological boundaries (including remnants of older compounds), vowel ablaut, contact influence, sporadic assimilation, and transcription differences.

2.2.2.1 Irregularity in Vowel Quality

The Influence of the Glottal Stop ʔ

The glottal stop **ʔ** influences the quality of the preceding vowel. The examples below illustrate variation in vowel reflexes across multiple languages for stems containing **ʔ**. In each set, the vowel reflex for each language is marked with (x) in the appropriate column. The first three examples are [431] **raʔ-* 'ripen, be ripe', [30] **kaʔ-* (PNHEC **kaʔ-*, **keʔ-*) 'to get up, arise', and [34] **baʔ-* 'be lost':

	[431]	**a**	**e**	**i**	[30]	**a**	**e**	**i**	[34]	**a**	**e**	**i**
PHEC	***raʔ-**				***kaʔ-**				***baʔ-**			
Burji		x				x					x	
Had				x				x				x
Kam			x				x			x		
Sid		x	x			x				x		
Ged		x				x				x		

For each of the above sets, there are three retentions of PHEC ***a**. The reflexes of Gedeo and Sidaama are the most stable across the three examples. Burji and Kambaata each reflect a regular **a** once when compared to the EC parallels. Hadiyya shows fronting to **i** in all three examples. Kambaata and Sidaama show **e** in the first two sets. Such diversity of reflexes, including three different vowels, suggests that the conditioning effect of **ʔ** was not uniform across HEC.

The second set includes [293] **waʔ-* 'water', [269] **t'aʔ-*, **t'eʔ-* (PNHEC **t'aʔ-*, **t'eʔ-*) 'be extinguished, to diminish', and [144] **meʔ-* 'how many? how much?':

	[293]	**a**	**o**	[269]	**a**	**e**	[144]	**e**	**i**
PHEC	***waʔ-**			***t'aʔ-, *t'eʔ-**			***meʔ-**		
Burji		x			x	x			x
Had			x		x			x	
Kam			x		x	x			
Sid		x				x		x	
Ged		x				x		x	

The alternations are limited to only two vowels in these examples. Again, Sidaama and Gedeo show uniformity throughout, while the rest show variation between **a**, **o**, **e**, and **i**. In [293], the rounded **o** in Hadiyya and Kambaata is likely due to the preceding labial **w**, with further incentives from the variation-triggering **ʔ**.

According to the distribution patterns, the attested vowels are chiefly the low vowel **a** and the front vowels **i** and **e**. Burji predominantly uses **a** (in 4 stems), with some **e** (2 stems), indicating a preference for low/mid vowels. Hadiyya uniquely prefers the high front vowel **i** (in 3 stems), with some **e**, **a**, and **o** (1 stem for each). Comparable to Hadiyya, Kambaata also has varied vowel reflexes, using **a** (2 stems) and **e** (3 stems), with **o** (1 stem). Sidaama uses both **a** (4 stems) and **e** (3 stems) reasonably equally, with doublets appearing in roots like **raʔ-* (both **a** and **e** forms). Gedeo consistently uses **a** (4 stems) and **e** (2 stems).

The double vowels, such as in PHEC [749] **yaaʔ-* 'assembly, meeting', PHEC [354] **yaaʔ-* 'to flow, assemble', and PHEC [168] **heeʔ-* 'to exist, to live', are not influenced by the glottal stop. However, exceptions to this generalization can be found in [440] **weeʔn-* 'Colobus monkey'. Overall, the variation of a single vowel may serve as additional supporting evidence for reconstructing a root or stem with a final glottal stop, in conjunction with other evidence.[72] Other factors to consider are assimilation in Hadiyya and the influence of the inserted **-y-** in Burji, which contribute to this variation. Additionally, there is a merger of the reflexes of PEC ***ʕ** with **ʔ** in HEC, where the earlier form is also known to condition vowel changes in some EC languages, e.g., Western Omo-Tana. For instance, the parallel of PHEC [431] in Afar is *laʕallaʕe* 'cook, fry, grill', *alaye* 'become ripe, ripen, cook', and the PEC reconstructions of [30] is **kaʕ-* 'get up, arise' and [117] is **ʃaʕ-* / **saʕ* 'cow'.

The only existing study on Cushitic that has proposed glottal stop influence is Hetzron (1980, p. 107). He discusses its phonetic effect on the following vowel in pronominal prefixes.

[72] A counterexample would be the variation found in the loanword: PHEC [341] **loom-* 'lemon' → Burji *loomay*; PNHEC **loom-* 'lemon' → Had *loomeʔe*, Kam *loommi-ta*, Sid, Ged *loome* (see Oro *loomii* and Amharic *lomi* 'lemon') unless one accepts that it has been borrowed at the proto-level and used to have a glottal stop or another elided final consonant. The same applies to [365] *bärbäre* 'ground pepper' from Amharic.

The Influence of the Labials b and w

Vowels before the labials **b** and **w** also exhibit some irregular variation in quality. The reflexes are listed below based on whether the vowel reflex is rounded or unrounded:

[4]	rounded	unround.		[74] rounded	unround.	
PHEC	***gulib-**	***gilub-**	'knee'	***k'ub-**	**-**	'fall (down)'
Burji	**-**	**(gilba)**	'knee'	**-**	**-**	-
Had	**gurubbo**	**-**	'knee'	**ub-**	**-**	'fall (down)'
Kam	**gulubi-**	**-**	'knee'	**ub-**	**-**	'fall (down)'
Sid	**gulube**	**-**	'knee'	**ub-**	**-**	'fall (down)'
Ged	**gulubo**	**-**	'knee'	**-**	**ib-**	'fall (down)'

[495]	**rounded**	**unround.**	
PHEC	**-**	***lab-**	'many, big, much'
Burji	**-**	**laboo**	'many, much'
Had	**lob**	**-**	'many, big, important
Kam	**-**	**(ligiba)**	'many'
Sid	**lobo**	**-**	'big, many, loud, much'
Ged	**(lumolla)**	**-**	'big'

[466]	**rounded**	**unround.**	
PHEC		***diwh-, *diwk-**	'empty'
Burji	**duwwa**	**-**	'empty and deserted house'
Had	**-**	**diiha**	'empty'
Kam	**-**	**diiha**	'empty'
Ged	**duwwi**	**-**	'empty, bare, naked'

The following examples are from NHEC only with an initial **w** reflex of ***ħ** before ***u**:

[519]	**rounded**	**unround.**		[163] **rounded**	**unround.**	
PNHEC	**-**	***ħililll-**	‘smoke’	***ħur-**	**-**	‘young ox’
Had	**wuriira**	**wiriira**	‘smoke’	**woʔla**	**-**	‘calf’
Kam	**-**	**wiliili-**	‘smoke’	**worjaamu**	**-**	‘young bull’
Sid	**-**	**wiliile**	‘smoke’	**wor-aamo**	**war-aamo**	‘ox, bull’
Ged	**-**	**wiliille**	‘smoke’	**-**	**-**	-

[607]	**rounded**	**unround.**		[520] **rounded**	**unround.**	
PNHEC	***ħuym-s-**	**-**	‘to fill’	***ħuy ʔ-m-**	**-**	‘be full’
Had	**wonʃ-**	**wanʃ-**	‘fill’	**woʔm-**	**-**	‘be full’
Kam	**-**	**wiinʃ-**	‘fill’	**-**	**wiim-**	‘be full’
Sid	**wonʃ-**	**-**	‘fill’	**woʔm-**	**-**	‘be full’
Ged	**wonʃ-**	**wanʃ-**	‘fill’	**woʔm-**	**-**	‘be full’

Reflexes of these two stems are discussed in §4.2.5.7.

There is an example with double unrounded vowel reflexes, but where Gedeo still shows irregular reflexes before the labial **-b** in PNHEC [306] **y ebb-* ‘be(come) hot/warm’ → Hadiyya, Kambaata, and Sidaama *iibb-*, Gedeo *eebb-* ‘be(come) hot/warm’. Below are more examples with single regular vowels without any variation in the reflexes before the labial **-b**:

	[768]		[183]		[300]	
PHEC	***hab-**	‘to forget’	***gib-**	‘to refuse, hate’	***k’ub-**	‘finger(-ring)’
Burji	**hab-**	‘forget’	**jibb-**	‘to hate’	**k’ubee**	‘finger-ring’
PNHEC	***hab-**	‘to forget’	***gib-**	‘to refuse’	***k’ub-**	‘finger(-ring)’
Had	**-**	-	**gib-**	‘flee’	**k’ubaʔa**	‘finger-ring’
Kam	**hab-**	‘forget’	**gib-**	‘refuse’	**k’ubbata**	‘finger-ring’
Sid	**hab-**	‘forget’	**gib-**	‘refuse’	**k’ubb-e**	‘ring’
Ged	**-**	-	**gib-**	‘refuse’	**k’ube**	‘ring’

Therefore, the preceding examples show that labials do not exert a regular rounding effect on vowels in comparable environments. Forms such as **hab-* ‘to forget’ and **gib-* ‘to refuse, hate’ retain unrounded vowels before **b**, so rounded reflexes in other sets are better treated as sporadic rather than as the expected outcome of labial influence. Such sporadic rounding is attested in Burji and Gedeo [466], and Hadiyya and Sidaama [495]. In [4], however, PNHEC **gulub*-does not allow a confident choice between PHEC **gulib-* and **gilub-* ‘knee’.

Some variation in vowel quantity and quality could be attributed to vowel coalescence as a result of intervocalic consonant loss ([115] and [114]) or compounding ([257]).

	[115]		[114]		[257]	
PHEC	***k'ab y-t**	'be cold'	***debeʔ-**	'to belch'	***ke**	'you' (sg.obj)
Burji	**kabbee d'aa-**	'be cold'	**deemmeys-**	'to belch'	**ʃeé**	'you' (sg.obj)
PNHEC	***k'iid-**	'be cold'	***debeʔ-**	'to belch'	***ke**	'you' (sg.obj)
Had	**k'iid-**	'be cold'	**deeʔacc-**	'belch'	**keésẹ**	'you' (sg.obj)
Kam	**gid-**	'be cold'	**debeeʔl-**	'belch'	**kees**	'you' (sg.obj)
Sid	**k'iid-**	'be cold'	**debeeʔl-**	'belch'	**-**	-
Ged	**k'iid-**	'be cold'	**(deeʔefat-)**	'belch'	**-**	-

The consonant loss and coalescence evident in the verbalized PNHEC stem [115] **k'iid-* 'be cold' occurred via labial lenition and glide loss **k'aid* ← **k'awid* (§§2.1.3, 2.1.4). In [257], Kambaata and Hadiyya combine **ke* and **ise* into **keese*.

Compounding or Affixation

Prior studies attribute the fluctuation of the root vowel to fossilized ablaut in the prefix conjugation in East Cushitic (Sasse, 1982, pp. 42, 26, 30; Sasse, 1979, p. 5). Hetzron (1980, p. 39) also believes that some vowels in the prefix serve as tense markers in East Cushitic languages. He also cites Diakonoff (1965, p. 84), where the latter implies that the vocalic alternation originally occurred within the root, and it has been shifted to the prefix in Cushitic. Taking the previous wider perspective, one might want to see if the remnants of the old prefix or verb compounding system may have played any role in the verb-initial vowel ablaut in HEC.

	[306]		[514]		[223]	
PHEC	***y ebb-**	'be hot/warm'	***ebb-**	'to bring'	***eg-**	'to wait'
Burji	**d'aybad'**	'to be warm'	**amb-**	'bring'	**eeg-**	'to wait'
PNHEC	***y ebb-**	'be hot/warm'	***ebb-**	'to bring'	***eger**	'to wait'
Had	**iibb-**	'be hot/warm'	**eeb-**	'bring'	**eger-, egar-**	'wait (for)'
Kam	**iibb-**	'be hot/warm'	**eeb-**	'bring'	**eger-, agar-**	'wait (for)'
Sid	**iibb-**	'be hot/warm'	**ʔabbi**	'bring'	**ʔagar-**	'wait'
Ged	**eebb-**	'be hot/warm'	**-**	-	**egert-, eged'-**	'wait!, wait (for)'

As [306] and [514] demonstrate, potential labial influence is evident in these examples. However, remnants of support verbs might be a more plausible

explanation within the context of the current research. At least [306] consists of the support verbs **y* 'to say, do, be' (§4.2.4) attached to a following (nominal) stem.[73] Example [514], on the other hand, might be a remnant of a prefix verb similar to PHEC [62] **-al-* 'give birth, beget (of human)' and [449] **k'al-* 'give birth, beget (of animals)'. As mentioned previously (§2.2.1.1, PHEC *e), [223] could also be alternatively explained in terms of vowel assimilation (**e** → **a** in $\mathbf{V_1}$) if one discards the prefixing possibility.

Numerals

The vowels and consonants of the numerals are mostly unstable, sometimes with more than one variant recorded in the same source. This is expected because they are the most susceptible to use in contact situations, and their variants are comparable to those found in other languages in the area.

	[38]		[19]		[142]	
PHEC	***liħ-**	'six'	***sazhet**	'eight'	***tam-n-**	'ten'
Burji	**líya, lia**	'six'	**hiditta**	'eight'	**tanna**	'ten'
PNHEC	***leħ-**	'six'	***sazzeett-**	'eight'	***tom-n-**	'ten'
Had	**loho**	'six'	**sadeento**	'eight'	**tommo**	'ten'
Kam	**lehu**	'six'	**hezzeettu**	'eight'	**torduma**	'ten'
Sid	**lee, leye**	'six'	**sette**	'eight'	**tonne**	'ten'
Ged	**-**	-	**saddeeta**	'eight'	**tomme**	'ten'

Similar to [142], Burji has a different vowel in PHEC [636] **omt-*, **om~omt-* 'five' → Burji *umutta* 'five'; PNHEC **ont-* 'five' → Hadiyya *onto* 'five', Kambaata *ontu* 'five', Sidaama *onte, ʔoonte* 'five', and Gedeo *onde* 'five'.

Some of the above examples include one irregular reflex in a single language, but there are more complex cases.

> PHEC [51] **lokk-, *lekk-* 'foot, leg' → Burji *lukka* 'foot, leg'; PNHEC **lokk-*, **lekk-* 'foot, leg' → Had *lokko* 'foot, leg', Kam *lokka-ta*, Sid *lekka*, Ged *lekka* 'foot, leg'

> PHEC [206] **awd-* 'threshing floor' → Burji *oyda*, *ooyda* 'threshing floor'; PNHEC **awd-* 'threshing floor' → Had *awda*, *oodoʔo*, Kam *oodu*, Sid *ʔaiddumme*, *ayddumme*, Ged *oodde* 'threshing floor'

[73] The support verb is placed before the verb as a remnant of the older order (i.e., when affixes used to occur before the root, so too did the support verbs, e.g., Gedeo *mar-* 'go' → *imar-* 'take' and *dag-* 'come' → *idag-* 'bring').

Three different corresponding vowel reflexes (**u**, **o**, **e**) are found in [51], and the vowel and glide sequences **oy**, **ooy**, **ai**, **aw**, and **oo** occur in [206]. Therefore, two proto-forms are reconstructed for [51] while the **u** of the Burji form is attributed to Oromo influence since the latter has *luka* 'foot, leg'. [206] **awd-* 'threshing floor' is more problematic, and thus is only provisionally reconstructed.

According to previous studies, vowel variations have been reported to include vowel laxing in closed syllables and word-internally in the HEC languages, with the exclusion of Sidaama. This laxing most commonly affects **a** and **i**, less commonly **e**, and rarely **o** and **u**. The lax allophones are **ə**, **ɪ**, **ɛ**, **ɔ**, and **ʊ**, respectively (Hudson, 1976, p. 249). For Burji, Sasse (1982, p. 16) describes these as vowel allophones that apply in cases of both single and double vowels. He also notes that the contrast between **a** and **e** is neutralized in closed syllables before **y** (see §2.2.1.2, PHEC *ee).

2.2.2.2 Irregularity in Vowel Quantity

There are several examples with irregularity in the quantity of vowels (double vowels corresponding to single ones), and in a few cases, quality as well. Where related derived stems exist, one may tentatively assume remnants or innovated morphophonological processes triggered by derivational morphology and analogy.

The first example includes the words PHEC [184] **reħ ʔ* 'to die', [439] **reħ-* 'death', and [329] **reħ-ʃ-* 'corpse', which share the same root originally.

	Verb		**Noun**		**Noun**	
Burji	**rey-, re-**	'die'	**reya, reeya, reeha**	'death'	**reesi, reefa**	'corpse'
Had	**leh-**	'die'	**leho**	'death'	**leeʃa, leeʃʃa**	'corpse'
Kam	**reh-**	'die'	**rehu-ta**	'death'	**re(e)ʃa**	'corpse'
Sid	**re-, reʔ-**	'die'	**reo, rewo**	'death'	**reeʃʃa, reeʃa**	'corpse'
Ged	**rey-, re(ʔ)-**	'die'	**reyo, reo**	'death'	**reenʃa, renʃa**	'corpse'

The parentheses around the Kambaata and Gedeo sounds indicate optional segments, e.g., in Kambaata, variants with single and double vowels have been recorded for both the noun and the verb.

Other sets derived from a common root include PNHEC [338] **yaʔ-* ‘to call’,[74] and PHEC [749] **yaaʔ-* ‘assembly, meeting’.

	Verb		**Noun**	
Burji	**-**	-	**yaa**	‘assembly, meeting’
Had	**yaaʔa, yaʔ-**	‘to call’	**yaʔ-**	‘large in mass’
Kam	**yaaʔ-**	‘hold (a meeting)’	**yaaʔa**	‘assembly of elders’
Sid	**yaaʔa**	‘come together’	**yaaʔa**	‘communal assembly’
Ged	**yeʔa**	‘to call’	**yaaʔa**	‘crowd’

The following is reconstructable for PHEC [168] **heeʔ-* ‘to exist’ and **heeʔ-ʃ-* ‘life, existence’ (see §4.1.2):

	Verb		**Noun**	
Had	**heeʔ-**	‘live, exist’	**heecca**	‘life, condition’
Kam	**heʔ-**	‘live’	**heecca-ta**	‘life, existence’
Sid	**heed’-**	‘exist, be alive’	**heeʃʃo**	‘life’
Ged	**hed’-**	‘exist, be there, live’	**heeʃʃo, hed’o**	‘life’

Some irregularity in quantity in the above examples could be due to the loss of the following glottal, for instance, in the derived nouns in **reħ-ʃ-* ‘corpse’ and **heeʔ-ʃ-* ‘life, existence’.

Variation in vowel quantity is also attested in the PHEC verb [548] **arraab-* ‘to lick’ and noun [9] **arrab-* ‘tongue’.

	Verb		**Noun**	
Burji	**arrab-, arraab-**	‘lick’	**arraaba**	‘tongue’
Had	**aall-**	‘lick’	**allabo**	‘tongue’
Kam	**arraab-**	‘lick’	**arrabi-ta**	‘tongue’
Sid	**-**	-	**ʔarrabo**	‘tongue’
Ged	**arraab-, arraap’p’-**	‘lick’	**arrabo**	‘tongue’

A more complex connection is proposed between the derived verbs PHEC [306] **y ebb-* → **d’ayb-* ‘be(come) hot/warm’ and [128] **t’eeb ʔ* ‘to be thirsty’ and the noun **t’eeb-* ‘thirst’.

[74] Set [354] with Burji *yaʔ-* ‘to flow’ and Gedeo *yaaʔa* ‘flow’ could also be argued to be cognates.

	Verb		**Verb**		**Noun**	
Burji	**d'aybad'**	'be hot/warm'	**d'aybood'-**	'become thirsty,	**d'ayboo**	'thirst'
Had	**iibb-**	'be hot/warm'	**t'eebeʔ-**	'be thirsty'	**t'eebo**	'thirst'
Kam	**iibb-**	'be hot/warm'	**t'eebek-, t'eebeʔ-**	'be very thirsty'	**t'eebeʔu**	'thirst'
Sid	**iibb-**	'be hot/warm'	**-**	-	**-**	-
Ged	**eebb-**	'be hot/warm'	**d'eeboo?-, d'eebot-**	'be thirsty'	**d'eebo**	'thirst'

A less clear semantic link could be assumed for PHEC [287] **t'ib- (ʔ)* 'to make trouble', [250] **t'iib-* 'to push', and [417] **t'ib-* 'disease, evil'.

	Verb		**Verb**		**Noun**	
Burji	**d'ib-**	'to make trouble'	**d'iib-**	'to push'	**d'iba**	'evil'
Had	**t'ib-**	'to hide'	**t'iib-**	'to push'	**t'isso**	'demon, evil'
Kam	**-**	-	**t'iib-**	'press'	**t'ida-ta**	'illness'
Sid	**d'ib-**	'be troubled'	**t'iib-**	'to push'	**t'ibba**	'disease'
Ged	**d'ib-**	'cause trouble'	**d'iib-**	'press, push'	**d'ibe**	'sickness'

One final example with three stems includes PHEC [252] **hasaaw-* (← **haʃaaw-*) 'speak, talk', [229] **haʃ~haʃ-* 'to whisper', and **haas-* 'speech'.

	Verb		**Verb**		**Noun**	
Burji	**haasaaw-**[75]	'speak, talk'	**hasaas-**	'whisper'	**hasaasaa**	'whisper'
Had	**hasaaw-**	'chat'	**haʃaʃa wac-**	'whisper'	**(wac'a)**	'speech'
Kam	**haasaaww-**	'discuss'	**haʃaaʃʃ-**	'whisper'	**haasaawwa**	'speech'
Sid	**hasaaw-**	'speak, talk'	**haʃaaʃ-**	'whisper'	**-**	-
Ged	**haaso?**	'speak, chat'	**haʃaʃ-**	'whisper'	**haaso?a**	'speech'

The above examples demonstrate that variation in vowel quantity can sometimes be the only difference between nouns and verbs, or derived verbs, with the verb having a single vowel against a double vowel in the noun, or vice versa, with or without additional suffixes. The other factor that might have contributed to the alternation in quantity is accent shift between the ultimate and penultimate syllables, since it is governed by syllable weight (§2.2). Therefore, this depends on the attached suffix and the subsequent degree of apocope that occurred in each word.

[75] The first syllable has a single vowel in *hasaawa* 'to talk' as transcribed in Getahun Addo (2017, p. 426). The dictionary transcribes the two verbs and the noun with a single vowel.

Here are pairs of nouns that are clearly semantically comparable, which could have led to the irregularity in their vowel correspondences through analogy:

	Noun		**Noun**	
Burji	**meena**	'people, man, person'	**meena**	'people'
Had	**mán-cọ**	'man, person'	**meent-íccọ, mento**	'woman'
Kam	**man-cu**	'man, woman'	**meent-iccu-ta**	'woman'
Sid	**man-co**	'man, woman'	**man-co**	'man, woman'
Ged	**manjo, manna**	'man, person'	**manjo, manjicco**	'woman'

They are reconstructed as PHEC [138] **man-ʃ-* (pl) 'person' and [215] **meen-t-* 'woman, girl' (see [609] **man~n-* 'people'). Derivation has contributed to the variation, and paradigmatic drift is responsible for the identical lexemes.

	Noun		**Noun**	
Burji	**c'iid'aa, c'iid'd'aa**	'bird'	**c'iwʔe, c'iwʔeena**	'chicken'
Had	**c'iiʔ-icco**	'small bird'	**c'iic'oolla, c'iic'oora**	'chick'
Kam	**c'iiʔ-iccu-ta, c'iiccu-ta**	'bird'	**c'iic'oorra**	'chick'
Sid	**c'eʔicco, c'eeʔ-icco**	'bird'	**c'aac'c'urre**	'chick'
Ged	**c'iicco, c'iʔa**	'bird'	**c'aac'ute**	'chicken'

The two nouns above are reconstructed as PHEC [348] **c'iiʔ-* 'bird' and [226] **c'iwʔ-* 'chick(en)'. According to Gasparini (1983, p. 56), Sidaama *c'eʔicco* is contractable to *c'îcco* 'bird'. Anbessa Teferra (2000, p. 27) explains that vowel deletion occurs when two different vowels occur in a sequence in Sidaama. However, Gasparini transcribes the remaining vowel **i** with a falling accent, indicating that its underlying quantity is double.

The above categorization should be taken as an arguable and provisional approximation with possible explanations. The variation in vowel quality and quantity requires further research to establish potential regularities or rules.

2.2.3 Terminal Vowels

Highland East Cushitic nominal stems end in vowels, referred to here as terminal vowels (TV). This section examines the distribution and variation of terminal vowels across the five HEC languages. Section 2.2.3.1 discusses identical TVs where cognates share the same final vowel quality. Section 2.2.3.2 examines cases of TV variation where cognate sets show different vowel qualities.

According to Hudson (1976, p. 249), HEC (Burji (Tesfaye Baye, 2015, p. 59) and Gedeo (Eyob Kelemework, 2015, p. 267)) words end in vowels except in Hadiyya and in a few instances in Kambaata. Such vowels in final position, i.e., terminal vowels (TV), are only reconstructed on the pronominals as arguably fossilized nominal suffixes (§4.1.5.2). The terminal vowels of the nouns may have originated from fossilized suffixes, recent independent suffixation, apocope of stem-final consonants, or coalescence with other suffixes.[76] Other than the arguably fossilized nominal suffixes, the alternative origins fall outside the scope of this analysis. Thus, terminal vowels are hyphenated, suggesting a morpheme boundary following the root final consonant.

Terminal vowels are widely reported as devoiced or extra-short in Burji (Sasse, 1982, p. 17; Tesfaye Baye, 2015, p. 46), Hadiyya (Tadesse Sibamo, 2015, p. 27), Kambaata (Treis, 2008, p. 20), and Gedeo (Eyob Kelemework, 2015, p. 42), while Sidaama appears more resistant to this reduction. Hudson (1976, p. 250) notes that devoicing is especially common after strong (or heavy) syllables and may render terminal **i** and **a** nearly inaudible; he also observes that final vowels may be lost altogether in continuous speech in Hadiyya, and less commonly in Kambaata.

2.2.3.1 Identical Terminal Vowels

The HEC languages show instances of identical corresponding terminal vowels. Only three vowels, **a**, **e**, and **o**, are commonly attested as terminal vowels. As terminal vowels, **i** and, to a lesser extent, **u** are only found on pronouns.

Terminal vowels are chiefly single vowels. However, there are instances of double terminal vowels, some of which are due to differences in transcription. For example, where Roba Dame & Wedekind (2008) have double terminal vowels, Hudson (1989) has a single terminal vowel in Burji data. The history of stress placement (§2.4) and contact has also contributed to such an anomaly.

[76] Underived verb roots typically end in consonants. Overall, the morphophonemic and morphological evidence shows that the PHEC roots must have ended in consonant(s), and those with vowels are their stems (including apocopated ones).

Terminal -i

Terminal **-i** is found on pronominals only.

	[2]		[27]		[6]	
PHEC	***an-i**	‘I’ (subj)	***at-i**	‘you’ (sg.subj)	***is-i**	‘she’ (subj)
Burji	**ʔánị**	‘I’ (subj)	**aʃi**	‘you’ (sg.subj)	**ʔíʃị**	‘she’ (subj)
PNHEC	***ani**	‘I’ (subj)	***ati**	‘you’ (sg.subj)	***isi**	‘she’ (subj)
Had	**ʔánị**	‘I’ (subj)	**ʔátị**	‘you’ (sg.subj)	**ʔísị**	‘she’ (subj)
Kam	**án**	‘I’ (subj)	**át**	‘you’ (sg.subj)	**íse**	‘she’ (subj)
Sid	**ʔani**	‘I’ (subj)	**ʔáti**	‘you’ (sg.subj)	**ise**	‘she’ (subj)
Ged	**ani**	‘I’ (subj)	**ati**	‘you’ (sg.subj)	**ise**	‘she’ (subj)

For sets [2] and [27], Hudson (1989) transcribes Kambaata *ani* ‘I’ and *ati* ‘you’ with the terminal vowels, while *án* ‘I’ and *át* ‘you’ from Treis (2008) have devoiced and inaudible terminal vowels. In [6], the irregular terminal vowel of Kambaata, Sidaama, and Gedeo *ise* ‘she’ is probably due to analogy with the accusative form of the pronoun *ise* ‘her’ (obj).

The demonstratives provide additional evidence of these terminal vowels (§4.1.5.1).

Terminal -e

Terminal **-e** is the least attested final vowel in comparison to final **-a** and **-o**. This section includes most of the examples where terminal **-e** is attested in the available data.

	[151]		[28]		[240]	
PHEC	***is-e**	‘her’	***ay-e**	‘who?’	***k’ort’umm-**	‘fish’
Burji	**ʔiʃeé**	‘her (obj)’	**ayyee**	‘who?’	**k’urt’ummee**	‘fish’
PNHEC	***ise**	‘her’	***aye**	‘who?’	***k’ort’umm-**	‘fish ’
Had	**ʔíse̩**	‘her (obj)’	**ayye**	‘who?’	**k’urt’umeʔe**	‘fish’
Kam	**isé-ta**	‘her (obj)’	**aye**	‘who?’	**k’urt’ummee-ta**	‘fish’
Sid	**ʔisé**	‘her (obj)’	**aye**	‘who?’	**k’ult’uʔme**	‘fish’
Ged	**isee**	‘her (obj)’	**ayye**	‘who?’	**k’ult’uʔme**	‘fish’

The terminal vowel **-e** is only arguably reconstructed for the pronominals, as a fossilized suffix. It is a feminine accusative ***-e** in [151] **is-e* ‘her’ and [28] **ay-e* ‘who?’. The nominative counterparts are [6] **is-i* ‘she’ and [28] **ay-i* ‘who?’ (→ Kambaata and Sidaama *ayi* (nom) ‘who?’).

Further examples with identical terminal **-e** attested in the HEC languages are [273] **biin-* 'mosquito, fly (sp.)', [478] **kott-* 'hoof, foot', [91] **mooyy-* 'mortar', [169] **k'unc'-* 'bark (of tree)', [264] **arr-* 'grey hair', and [440] **wee?n-* 'Colobus monkey'.

Some instances are only identical in NHEC, excluding Kambaata:

> PNHEC [1] **ill-* 'eye' → Had, Sid, Ged *ille* 'eye' (see Kam *illi-ta* 'eye, seed, grain' and Burji *illa* 'eye').
>
> PNHEC [544] **c'ub-* 'dagger, knife' → Had *c'ube?e*, Sid *c'uube*, Ged *c'uup'e* (see Kam *c'up'i-ta* 'dagger').
>
> PNHEC [271] **ar-* 'wife' → Had, Sid, Ged *are* 'wife', (see Kam *ari-ta* 'wife', PNHEC [212] **ar-o* 'husband').

As shown in the above examples, nouns are always reconstructed without the terminal vowel.

Some loanwords have identical terminal **-e** vowels. for instance, [240] **k'ort'umm-* 'fish' from Oromo *k'urt'ummii* 'fish' (Sasse, 1982, p. 129) and other sets listed above. More loanwords are found with the terminal vowel **-a**.

Terminal -a

The terminal vowel **-a** is the most frequent in HEC, and the following are a few examples:

	[67]		**[68]**		**[167]**	
PHEC	***ang-**	'hand, arm'	***malaħ-**	'pus'	***fayy-**	'healthy'
Burji	**anga, ganaa**	'palm, hand'	**malaa**	'pus'	**fayya-n-ga**	'healthy'
PNHEC	***ang-**	'hand, arm'	***malaħ-**	'pus'	***fayy-**	'healthy'
Had	**anga**	'hand, arm'	**mara**	'pus'	**fayya?a**	'healthy'
Kam	**anga-ta**	'hand, arm'	**mala-ta**	'pus'	**fayya**	'healthy'
Sid	**?anga**	'hand, arm'	**malaa**	'pus'	**fayyaaleessa**	'healthy'
Ged	**anga**	'hand, arm'	**mala**	'pus'	**fayya**	'healthy'

The above examples show the addition of suffixes following the terminal vowel: **-ta** in Kambaata, **-n-ga** in Burji, and **-a** with epenthetic **-ʔ** in Hadiyya (§3.3.4). The reflexes of PHEC [68] **malaħ-* 'pus' indicate that some identical terminal vowels are a result of apocope.

Further examples with identical terminal **-a** vowels are found in the individual languages on the following PHEC stems: [66] **anuun-* 'breast', [126] **gog-* 'skin, hide', [383] **orz-* 'hair of animals', [209] **laaf-* 'soft, weak', [284] **ayyaan-* 'festival, holiday', [462] **meet-* 'scraping board for use in preparing ensete', [547] **am-* 'mother', and [609] **man~n-* 'people'.

Some instances of terminal **-a** might be part of (fossilized) plurative marking (§4.1.1.2) where there is reduplication *C_i~C_i-a, e.g., PHEC [609] **man~n-* 'people' (see [138] **man-ʃ-* 'person').

Similar to **-e**, numerous loanwords have identical terminal **-a** vowels. At least in Kambaata, Treis (p.c.) notes that the consonant-final Amharic loanwords receive a default terminal **-a**.[77]

> Unknown source: Burji [349] *gabalaa* 'limit', Had, Kam, Sid, Ged *gabala* 'border, limit'.
>
> Amharic *atär* 'peas': Burji [283] *ataraa* 'peas', Had, Kam, Sid, Ged *atara* 'peas'.
>
> Amharic *ʃəmbəra* 'chick-peas': Burji, Had, Sid, Ged [611] *ʃimbura* 'chick-peas', Kam *ʃumburu*, *ʃumbura* 'chick-peas' (see Oro *ʃumburaa* 'chick-peas').

The source of [349] is unknown, but it is identified as a loanword because of the irregular **l** in Hadiyya in place of the expected **r** reflex (§2.1.1.3, PHEC *l). The direction of transfer of the last two examples is debatable. Additional evidence of loanword status is the retention of intervocalic **b**, which should have been lenited in intervocalic position (§2.1.3).

Hudson (1989, p. 12) notes that in all five languages, though less often in Sidaama, the unstressed **-a** is pronounced as [ə]. This observation is relevant for the many instances of reduction of vowels in terminal position in HEC, e.g., inaudible Kambaata **-i** in [2] *án* 'I' and [27] *át* 'you' discussed above. The stress is obviously on $\mathbf{V_1}$ in the two examples, while the unstressed terminal vowels are reduced to zero (see §4.1.5.2 (Nominative)).

[77] Two more examples are in sets [325] Amharic *dida*, *dəda* 'mute', Oromo *duudaa* 'mute' and [350] Amharic *sänga*, Oromo *sangaa* 'gelding, bullock, ox'.

Terminal -o

Identical terminal **-o** vowels are most frequently attested in NHEC, where Burji has a different vowel. It regularly corresponds to **-u** in Kambaata with a few exceptions. The following are some rare examples, including Burji.

	[407]		[198]		[707]	
PHEC	***ebel-, *ebal-**	'so-and-so'	***tum-am-ʃ-, *tum-t-i-ʃʃ-, *tum-t-**	'blacksmith '	***magan-**	'sky, God'
Burji	**ebeloo**	'so-and-so'	**tuntoo**	'blacksmith'	**maganoo**	'mist'
PNHEC	***ebel-, *ebal-**	'so-and-so'	***tum-am-ʃ-, *tum-t-i-ʃʃ-, *tum-t-**	'blacksmith '	***magan-**	'sky, God '
Had	**ebaro**	'so-and-so'	**tumaanco**	'blacksmith'	**-**	-
Kam	**ebalo**	'so-and-so'	**tumaan-cu**	'blacksmith'	**maganu**	'God'
Sid	**ʔebelo**	'so-and-so'	**tunto**	'blacksmith'	**magano**	'God, sky'
Ged	**ebelo**	'so-and-so'	**tunto**	'blacksmith'	**mageno**	'God, sky'

In set [407], **ebel-, *ebal-* 'so-and-so' behaves similarly to pronouns with TV **-o** and **-e** in Hadiyya and Kambaata. In other words, they show gender distinction between the masculine accusative Hadiyya *ebaro*, Kambaata *ebalo,* and the feminine accusative Hadiyya *ebare*, Kambaata *ebale* 'so-and-so'. Kambaata **-o** is an irregular reflex, as the expected vowel is **-u**. The masculine counterpart **-o** of the feminine accusative **-e** above is also inferred to be an arguably fossilized suffix (§4.1.5.2) from the pronoun in [292]:

> PHEC [292] **is-o* 'him' → Burji *ʔisi̯*; PNHEC **is-o* 'him' → Kam *isú*, Sid *ʔisó*, Ged *isoo* (see Had *it't'o* 'him').

Kambaata *isú* 'him' shows a regular **-u**. In Burji, the vowel **-o** of the object pronoun has likely been restructured in analogy with that of the subject pronouns [7] *ʔísi̯* 'he, it' and [6] *ʔíʃi̯* 'she' and became *ʔisi̯* 'him'.

Further examples with identical final **-o** in the HEC languages include: PHEC [235] **od-* 'news', [477] **hobol-* 'boat', PNHEC [90] **malab-* 'honey', [70] **afoʔ-* 'mouth, opening, language', [14] **az-* 'milk', [444] **bagaz-* 'spear', [439] **reħ-* 'death', [276] **ball-* 'brother-in-law, father-in-law'. Burji's non-identical terminal vowels corresponding to NHEC **-o** in these examples are either **-a** or **-i**.

The HEC singulative suffixes receive a terminal vowel **-o** (§4.1.1.1) or the regular Kambaata reflex **-u**.

> Had, Sid [65] *faraʃʃo*, Kam *faaʃʃu*, Ged *faracco* 'horse'.

Had [138] *mán-cọ* ‘man, person’, Kam *man-cu* ‘man’, *man-cu-ta* ‘woman’, Sid *man-co* ‘man, woman, man, person’, Ged *manjo* ‘man, person’.

Burji [498] *woccoo*, Had *waʃicco*, Kam *woʃ-iccu*, Sid *woʃʃicco*, Ged *wolco* ‘dog’.

Had [602] *kin-c* ‘stone, rock’, Sid *kin-co* ‘stone, rock’, Ged *kin-jo* ‘stone, rock’.

Had [602] *kin-c* ‘stone, rock’ is another instance of terminal vowel reduction, which is quite common in Hadiyya.

Concerning Kambaata **-u**, there are examples where Treis (2008) and Treis (p.c.) have TV **-u** while Hudson (1989) sometimes transcribes TV **-o** and vice versa.

	Treis (2008; p.c.)	Hudson (1989)	
[230]	**t’abaru**	**t’abaro**	‘ashes’
[144]	**meʔu**	**meʔo**	‘how many? how much?’

Such transcription variation does not exist in the TV **-o** of the singulative, which is consistently transcribed as **-u**.

Finally, there are no correspondences with identical terminal **-u** in HEC. Terminal **-u** is mainly found in Kambaata, corresponding to **-o** in the rest of HEC, as mentioned above. Treis (p.c.) notes that there are individual instances of alternating **-u** ~ **-i** variants in Kambaata nouns.

2.2.3.2 Variation in Terminal Vowels

There are several reflexes that show variation in vowel quality across the individual HEC languages. The following examples include all HEC languages, with none having identical TV, although some of the regular correspondences discussed above occur.

	PHEC	Gloss	Had	Ged	Sid	Kam	Burji
[174]	*hamas-	‘snake’	-aʃʃa	-me	-o	-u	-i
[137]	*lal-	‘cattle’	-o	-o, -i	-o	-u	-i
[444]	*bagaz-	‘spear’	-o	-o	-o	-u	-i
[95]	*ʃomb-	‘lung’	-	-o	-o	-	-i
[606]	*gubeed-	‘thigh’	-o	-a	-a	-	-i
[9]	*arrab-	‘tongue’	-o	-o	-o	-i-ta	-a
[4]	*gulib-	‘knee’	*-co	-o	-co, -e	-i-ta	-a
[382]	*oll-	‘neighbour’	-aʔa	-aʔa	-a	-oo	-a
[51]	*lokk-, *lekk-	‘foot, leg’	-o	-a	-a	-a-ta	-a
[293]	*waʔ-	‘water’	-o	-aʔe, -eʔe	-a	-a	-a
[88]	*gudum-	‘shoulder’	-o	-o	-	-u	-a
[195]	*fal-, *falfal-	‘omen, magic’	-o	-o	-	-a	-aa
[5]	*san-	‘nose’	-e	-o	-o	-u-ta	-a
[439]	*reħ-	‘death’	-o	-o	-o	-u-ta	-a
[206]	*awd-	‘threshing floor’	-a	-e	-umme	-u	-a
[408]	*mut-	‘awl for basketwork’	-a	-a	-C’C’e	-a	-a
[381]	*k’ul-, *k’ul~k’ul~l-ʔ-	‘leech’	-a	-a	-e	-a	-ee
[1]	*ill-	‘eye’	-e	-e	-e	-i-ta	-a
[415]	*afal-	‘liver’	-e	-	-e	-i-ta	-a
[54]	*min-	‘house’	-e	-e	-e	-i	-a
[550]	*ibb-	‘louse’	-a	-e	-e, -iicco	-i-ta	-a
[224]	*t’ur-	‘dirt of body, clothing’	-a	-e	-e	-i	-ee
[544]	*c’ub- (PNHEC)	‘dagger’	-eʔe	-e	-e, -a	-i-ta	-

The examples show how the terminal vowel varies. A dash “-” indicates that no cognate is attested in the available sources. Kambaata **-ta** is the feminine accusative suffix, and it was likely added after the TV (mostly **-i** and **-u**). Some of the examples include further suffixes, e.g., the singulative or derivation suffix in [174], [550], and [4][78]. In [408], Sidaama **-C’C’** includes a glottalized stem-final **-t’t’-** probably triggered by a glottal suffix (§3.3.1). In the examples where there is more than one TV for the same stem and the same language, e.g., Sidaama **-e, -a** in [544] and Gedeo **-o, -i** in [137], each is from a different source. They most likely include mistranscriptions, especially in the Gedeo example, because Gedeo nouns do not end in TV **-i**. Kambaata also has variant forms mistranscribed by Hudson (Treis, p.c.).

[78] Hadiyya [4] **-co** has assimilated to the stem-final **-b**, yielding a geminate in *gurubbo* ‘knee’. This might also have happened to other examples with a terminal **-o**.

Regardless of the variation, if one looks for patterns, Sidaama and Kambaata cognates show the highest number of systematically corresponding TVs, with this occurring in more than half of the above 23 examples.

	Sidaama	**Kambaata**
[224]	**-e**	**-i**
[415]	**-e**	**-i-ta**
[550]	**-e**	**-i-ta**
[1]	**-e**	**-i-ta**
[4]	**-e**	**-i-ta**
[54]	**-e**	**-i**
[174]	**-o**	**-u**
[444]	**-o**	**-u**
[137]	**-o**	**-u**
[206]	**-umme**	**-u**
[439]	**-o**	**-u-t**
[5]	**-o**	**-u-ta**
[293]	**-a**	**-a**
[51]	**-a**	**-a-ta**

Sidaama **-e**, **-o**, and **-a** correspond to Kambaata **-i**, **-u**, and **-a**, respectively.

The only TV that shows some degree of similarity across all HEC languages is **-a** in the following examples, with variant TVs indicated in parentheses:

	Hadiyya	Gedeo	Sidaama	Kambaata	Burji
[51]	**(-o)**	**-a**	**-a**	**-a-ta**	**-a**
[293]	**(-o)**	**-aʔe, (-eʔe)**	**-a**	**-a**	**-a**
[382]	**-aʔa**	**-aʔa**	**-a**	**(-oo)**	**-a**
[408]	**-a**	**-a**	**(-C'C'e)**	**-a**	**-a**
[381]	**-a**	**-a**	**(-e)**	**-a**	**(-ee)**

This level of TV **-a** similarity is expected since it is the most common among the identical terminal vowels (§2.2.3.1).

The most frequent terminal vowels in HEC are **-a**, **-o**, and **-e**.[79] Terminal **-u** is not attested as an identical TV in cognate sets (except demonstratives), and **-i** appears only on pronominals. The data reveal correspondence patterns between Burji and NHEC. Where Burji has **-a**, NHEC languages (Gedeo, Hadiyya, and Sidaama) often have a corresponding **-o**. Kambaata frequently has **-u** or **-i** corresponding to **-o** and **-e**, respectively, in the rest of NHEC. The terminal vowels are removed from the stem when the nominal stem is derived; therefore, they are not part of the root. However, since HEC nouns must end in a vowel, TVs fulfill an obligatory phonotactic requirement. The alternative hypothesis about the pre-fossilization functions of the terminal vowels and a discussion of the above examples of TV variation are detailed in §4.1.5.2.

2.2.4 Summary

PHEC had a five-vowel system with contrastive quantity: ***i**, ***e**, ***a**, ***o**, ***u** and their double counterparts ***ii**, ***ee**, ***aa**, ***oo**, ***uu**. This system is securely reconstructed for PHEC and PNHEC and remains stable across all five HEC languages.

The regular stem vowels show consistent reflexes in cognate sets (§2.2.1). Single vowels are well attested, while double vowels, though less frequent in the database, are equally secure. The vowel ***e** is relatively infrequent compared to other vowels, a pattern that persists across HEC.

Deviations from regular correspondences fall into two categories (§2.2.2). Quality irregularities often result from adjacent consonants, particularly glottal stops, which condition vowel raising or lowering before **ʔ**. Quantity irregularities, where single vowels correspond to double vowels, reflect morphophonological processes, contact influence (especially from Oromo), or differences in transcription across sources.

Terminal vowels on nominal stems also exhibit both systematic correspondences and variation (§2.2.3). The vowels **-a**, **-o**, and **-e** are the most frequent TVs, while **-i** is restricted to pronominals. Two systematic patterns characterize TV variation: (a) Burji **-i** regularly corresponds to NHEC **-o**, and (b) within NHEC, Sidaama **-e** corresponds to Kambaata **-i**, while

[79] The same order of frequency and absence of final **u** or **i** have been noted synchronically in Hadiyya (Tadesse Sibamo, 2015, p. 27) as has the lack of final **u** in nouns in Burji (Tesfaye Baye, 2015, p. 48).

Sidaama **-o** corresponds to Kambaata **-u**. Terminal vowels are devoiced in all HEC languages except Sidaama.

These findings support the reconstruction of a stable five-vowel system with quantity contrast for PHEC. The documented irregularities in vowel quality, quantity, and terminal position reflect processes such as phonetic conditioning, morphophonology, apocope, and fossilized suffixation (§4.1.5.2), and contact.

2.3 Syllable and Root Structure

This section provides an overview of the syllable structure in PHEC and PNHEC. It focuses specifically on the development of these structures into their reflexes within Burji, Hadiyya, Kambaata, Sidaama, and Gedeo.

I use “open syllable” for syllables ending in a vowel (CV, V) and “closed syllable” for syllables ending in a consonant (CVC, VC). Syllable weight is treated moraically: light syllables (V, CV) contrast with heavy syllables created by vowel doubling and/or codas (CVV, VVC, CVC, CVVC).

The reconstructed PHEC and PNHEC forms suggest a common syllable structure of CV, CVC, and CVVC types. This system is characterized by a preference for simple onsets and codas at the root level, but with notable instances of double vowels, geminate consonants, and limited heterogeneous consonant sequences, mostly due to morpheme boundary alternation processes.

Thus, PHEC syllable structure follows the template (C)V(V)(C), permitting both open and closed syllables. The following subsections examine onsets (§2.3.1), nuclei (§2.3.2), codas (§2.3.3), and common root structures (§2.3.4). Syllable weight and its relationship to accent are discussed in §2.4.

2.3.1 Syllable Onsets

The onset of syllables in both PHEC and PNHEC is a single consonant (C). There are no initial consonant clusters (e.g., CCV). Onsets can be categorized into single consonant and zero onset.

The onsets of most of the reconstructed stems occur with a single consonant, including plosives, fricatives, nasals, and liquids, for example:

	PHEC	Gloss
[40]	***tuf-**	'to spit'
[351]	***t'uut'-**	'to suck'
[117]	***saʔ-**	'cow'
[307]	***mit-**	'one'
[209]	***laaf-**	'soft, weak'
[304]	***wit'-**	'to sow'

Overall, the plosives and non-pulmonic consonants function as common onsets; thus, syllables with low sonority onsets are most frequent.

The seemingly vowel-initial words in the HEC languages often have complex histories that obscure their original onset. In reconstruction, I therefore allow zero onset for a subset of PHEC/PNHEC forms rather than positing a systematic word-initial ***ʔ** (§2.1.5.1, PHEC *ʔ). One significant sound change defining PNHEC is the loss of the old initial non-pulmonic velar ***k'** (§2.1.5.1, PHEC *k'). The reconstruction PHEC [29] **k'ab-* → PNHEC **ab-* 'to hold/seize' is the best example: Where Burji and other EC languages, such as Oromo, retain forms with **k'**, the PNHEC languages reflect a root **ab-*, leading to reflexes like Kambaata *ap'p'-*. This loss of a **C** onset systematically changed the syllable structure of a whole set of words from CVC to VC in the common ancestor of the Northern branch. Additionally, the weakening and loss of initial ***ħ** is another major source of vowel-initial words. Examples of vowel-initial words include:

	PHEC	Gloss
[206]	***awd-**	'threshing floor'
[298]	***al ʔ**	'to be useful'
[136]	***eel-**	'pond, pool'
[62]	***al-**	'give birth, beget (of human)'

These examples show reflexes with and without consonant-initial onset in EC, demonstrating the variable stability of this onset.

In summary, PHEC onsets consist of single consonants or zero (some vowel-initial forms derive from the loss of **k'** and **ħ** in PNHEC).

2.3.2 Syllable Nuclei

The nucleus of the syllable is a vowel. Both single (V) and double (VV) vowels are phonemic (§2.2.1) and represent the peak of the sonority in the syllable.

Single vowels are common in many reconstructions (§2.2.1.1), e.g.:

	PHEC	Gloss
[268]	***mur-**	‘to cut’
[547]	***am-**	‘mother’
[339]	***ful-**	‘to go out, exit’
[415]	***afal-**	‘liver’

The most common vowel nuclei are **a**, **u**, and **i**. Terminal vowels (§2.2.3) further affect surface syllable structure by adding final open syllables to otherwise closed roots.

Though not to the extent of single vowels, double vowels are also frequent (§2.2.1.2), creating heavy syllables such as in the following lexemes:

	PHEC	Gloss
[635]	***maal-**	‘meat’
[390]	***hiit’-**	‘to stretch’
[330]	***sooz-**	‘dawn’
[325]	***duud-**	‘mute (person)’

A small number of vowel+glide sequences and heterovocalic sequences (especially in Burji and Gedeo) are best treated as later developments; relevant examples are discussed in §2.2.1.2, PHEC *ee. Section 2.2.2.2 includes examples of irregularity in vowel quantity and potential involvement of morpheme boundary alternation processes and morphological derivation.

2.3.3 Syllable Codas

Syllables can be open (ending in a vowel, e.g., CV) or closed (ending in a consonant, e.g., CVC). When closed in reconstructed roots, the coda structure shows some variation. The codas include single consonants (C), geminate consonants (CC), and consonant clusters (cC).

Single consonant codas are the most common type of closed syllable at the reconstructed root level, e.g.:

	PHEC	Gloss	Single Codas
[5]	***san-**	‘nose’	CVC
[36]	***gam-**	‘to bite’	CVC
[38]	***liħ-**	‘six’	CVC
[136]	***eel-**	‘pond, pool’	VVC
[551]	***it-**	‘to eat’	VC

Forms with added vowel suffixes (e.g., case) are therefore treated separately. For example, pronominal [151] **is-e* ‘her’ reflects the root **is-* plus accusative

***-e** (§4.1.5.2). The root itself has the shape VC, but once the suffixal vowel is added, the final consonant is resyllabified as the onset of the following syllable, *i.se*. Such forms are therefore not counted here as closed root syllables.

The geminate (doubled) consonant codas (CC) are less frequent at the reconstructed root level but are still attested, e.g.:

	PHEC	**Gloss**	**Single/geminate Codas**
[9]	***ar.rab-**	'tongue'	VC.CVC
[703]	***kaakk y-s-**	'to cackle (hen)'	CVVCC-VC
[150]	***baall-**	'feather'	CVVCC
[167]	***fayy-**	'healthy'	CVCC
[447]	***c'ubb-**	'sin'	CVCC

However, many of these geminates might be products of earlier boundary alternation processes, discussed in §§3.1.1 and 4.1.1.2 (Reduplication *C_i~C_i-a). Certain cases of assimilation of the pronominal **-t** to the stem-final **C** (§3.1.1) produce geminates – an innovation in NHEC, supporting Burji's earlier split from the other HEC languages.

Root/stem-final consonant clusters (represented as cC) are rare. Such morpheme-final clusters appear particularly before derivational suffixes, as discussed in Chapters 3 (Alternations at Morpheme Boundaries) and 4 (Morphology), e.g., [215] **meen-t-* 'woman, girl'. Sometimes the same derivational suffix yields different morphophonological reflexes in individual languages: regressive/progressive assimilation, fusion, palatalization, metathesis, and epenthesis. However, there are a couple of words reconstructed with **cC** clusters where **c** is usually a sonorant, and **C** is an obstruent, such as:

	PHEC	**Gloss**	**Cluster Codas**
[67]	***ang-**	'hand, arm'	VcC
[85]	***arg-**	'to lend/borrow'	VcC
[95]	***ʃomb-**	'lung'	CVcC
[161]	***bark-**	'head-support, pillow (of wood)'	CVcC

Two metathesis processes affecting coda consonants are especially important here because they also serve as diagnostic innovations for subgrouping within HEC. Unique to Burji is the metathesis of a velar and a sibilant. The limited examples of this metathesis include the derivational **-s** following stems:

	PHEC	Intermediate stage	Burji	Gloss
[45]	***fiik' ʔ**	***fiik'-s-**	**fiisk'-**	'to whistle'
[701]	***gig-**	***gig-s-**	**jisk-**	'to demolish, destroy'

For [701], Burji also has an underived stem *jig-* 'to demolish/destroy'. In Burji, the **Cs** sequence metathesized to *fiisk'-* and *jisk-*, respectively. In the NHEC languages, however, this sequence evolved differently (e.g., Gedeo *fiiss-* 'to whistle', Hadiyya *fiink'ees-* 'whistle', showing assimilation and other changes). This stop-sibilant metathesis is a distinct innovation of the Burji line after its split from what would become PNHEC.

A compelling piece of evidence for the unity of the NHEC languages as a group separate from Burji is their shared treatment of stem-final sonorant plus glottal sequences.

	[285]		[170]		[120]	
PHEC	***balʔ-**	'broad, wide'	***sum ʔ**	'urinate'	***fil-ʔ-**	'to comb'
Burji	**balʔad'-**	'become wide, broad'	**sinʔ-aaw-**	'urinate'	**sil-, silʔ-**	'comb'
Ged	**baʔla**	'wide, broad'	**ʃiʔn-**	'urinate'	**fil-, fiʔl**	'comb'

The reconstruction for the above examples contains the sequence **cʔ** (see also [22]). Burji preserves this order faithfully with its reflexes **lʔ** and **nʔ**. However, the NHEC languages, such as Gedeo, show reflexes with the metathesized sequences **ʔl** and **ʔn**. These Burji and Gedeo reflexes are the most transparent examples of this shift. Such metathesis (discussed in §3.3.2) represents a shared innovation that must have occurred in the common ancestor of the NHEC languages after the split from Burji, supporting the status of PNHEC as a valid linguistic node.

In general, the complex consonant sequences discussed here are evaluated as part of reconstructed root structure, not as straightforward reflections of later surface word forms. Many clusters and geminates arise through suffixation or later phonological developments presented in the following chapter, and forms that later take a terminal vowel or a vowel suffix undergo resyllabification.

2.3.4 Common Word and Root Structures

Based on the analysis of onsets, nuclei, and codas, the following primary root structures can be identified for monosyllabic and disyllabic roots:

Monosyllabic	**Disyllabic**
(C)VC	(C)V.CVC
(C)VVC	(C)VC.CVC
(C)VCC	
(C)VcC	

The parentheses around the consonant (C) indicate the presence or absence of an initial consonant, and the period (.) between the disyllabic syllables marks the syllable boundary. The reconstructed roots typically end in consonants, yielding closed final syllables in the underlying structure. The closed final syllables in PHEC roots attract the accent (§2.4), while open syllables are restricted to initial position in disyllabic words or to affixes.[80]

Hudson's (1976) and Wedekind's (1980) designation of **CV** as the most frequent syllable structure reflects synchronic reality, i.e., nominals end in a terminal vowel by default (§2.2.3). This also explains Wedekind's basic syllable structure formula (-C)CV(V)(-), representing eight different syllable types found in Burji, Gedeo, and Sidaama. However, Wedekind's (1980, p. 141) claim that "no closed syllables have been postulated" is an interesting analytical stance. He argues that it simplifies the description of consonant clusters. In this model, a word like Burji *gilb-* 'kneel' is likely syllabified as *gi.lba*, avoiding a closed *gil*. syllable. A consonant like the **l** here might be considered "extra-syllabic" or part of a complex onset in the following syllable. However, this is an analytical choice for phonological modeling. For all practical purposes, CVC structures are indeed phonetically and functionally present. The above analysis shows reconstructable sonorant-obstruent clusters and explicitly acknowledges codas. Thus, the phonotactic reality includes word-medial consonants that close preceding syllables.

2.3.5 Summary

In summary, the reconstructed forms of both PHEC and PNHEC point to a phonotactic system that favors simple syllables like **CV** and **CVC**. The system is also rich with phonemic vowel doubling and various types of codas, which create heavy syllables (**CVV**, **CVVC**, **CVC**). Consonant geminates and clusters are avoided in word-initial position but can appear in the coda, especially at morpheme boundaries.

The change of PHEC reflexes in the daughter languages was driven by a series of phonological innovations related to consonant sequences. The

[80] There are rare cases of roots reconstructed with a single consonant or vowel only, including the support verbs **y* 'to say, do, be' and **ʔ* 'to do'.

primary split between Burji and the NHEC languages is fully reflected in their divergent rules for consonant clusters, while the unity of the NHEC languages is clearly demonstrated by shared innovations like the sonorant-glottal metathesis (§3.3.2) and loss of initial ***k'*** and ***ħ** (§2.1.5). Within HEC, further micro-variations, such as the direction of assimilation or the specifics of diphthongs in Gedeo, show the ongoing change of the system.

2.4 Accent

In Highland East Cushitic, accent (often realized as greater prominence and/or higher pitch) is most clearly observed in the distribution of prominence across the last two syllables of the word. Based on the comparative evidence available here, I propose that the historically conservative pattern targets the rightmost heavy syllable (VV or closed), with penultimate accent emerging in many modern forms through later reductions affecting final syllables.

In the various HEC languages, the discussion around accent (or stress) placement has been confined to the penultimate and ultimate syllables. It seems that the accent predominantly falls on the penultimate syllable synchronically. As mentioned above and illustrated below, this is likely a result of historical changes, such as the reduction of terminal double vowels and apocope.[81]

This section examines accent patterns in Highland East Cushitic languages, with a particular focus on the relationship between syllable weight, vowel quantity, and accent placement. The analysis demonstrates that the accent historically falls on the ultimate heavy syllable (that having a double vowel or being closed by a consonant), with systematic shifts to the penultimate in contexts of terminal vowel reduction. These patterns provide additional evidence for the moraic analysis of HEC vowel quantity.

[81] Data sources: Burji (Tesfaye Baye, 2015, pp. 63–64; Wedekind, 1980, p. 140), Hadiyya (Tadesse Sibamo, 2015, pp. 31–32), Sidaama (Kawachi, 2007, pp. 63–66), Gedeo (Eyob Kelemework, 2015, pp. 91–92; Wedekind, 1980, p. 140), Kambaata (Treis, 2008, pp. 52–53).

	Penultimate		**Ultimate**	
Burji	**k'ál̪a**	'child'	**sotteé**	'bowel'
	máal̪a	'meat'	**koyloó**	'cheek'
Hadiyya	**wóʔo**	'water'	**-**	-
	maára	'meat'	**-**	-
Sidaama	**búko**	'dust'	**mundée**	'blood'
	seéda	'long, tall'	**aróo**	'husband'
Gedeo	**béltọ**	'daughter' (subj)	**giirá**	'fire'
	báabbạ	'Mother!' (voc)	**biiʔnné**	'mosquito'
Kambaata	**sáʔu**	'cow' (m.nom)	**saʔá**	'cow' (m.acc)
	ʃóolit	'four' (f.nom)	**tassáa**	'heifer' (m.acc)

Hadiyya does not place an accent on the final syllable (Tadesse Sibamo, 2015, pp. 31–32; Dereje Adane, 2022, pp. 23–24) due to the significant reduction of TVs. As a result, the former penultimate syllable can be interpreted as the current final one when the last vowel is inaudible. In their discussions of stress and accent in Gedeo, neither Dawit Tilahun (2019, pp. 53–55) nor Eyob Kelemework (2015, pp. 91–95) provides examples of accentuation on the penultimate syllable of stems that are not extended with a suffix or in citation forms. However, Wedekind (1980, p. 140) does present examples and notes that certain forms of Gedeo nouns that he calls the non-objective cases and the singular imperative of the verb also have penultimate stress, even in the absence of an "affix".

Unlike Burji and Kambaata, the Hadiyya and Sidaama languages demonstrate a different placement of the accent within the second mora of bimoraic syllables. Kawachi's (2007, p. 64) analysis explains that in Sidaama, "when the penultimate syllable is a heavy syllable, the second vowel component of it is more prominent than the first one".[82] This observation likely applies to Hadiyya as well. For Kambaata, Treis (2008, p. 52) places the accent marker on the first mora, even though both morae in dimoraic syllables have equal prominence. This distinction also supports the terminology used in this study, where the terms "single" and "double" are employed to describe the vowels of HEC languages, emphasizing quantity based on morae rather than simply length.

[82] However, Kawachi (2007, p. 64), in disagreement with all his predecessors, claims that the high pitch accent occurs on the penultimate vowel segment, not on the penultimate syllable, as shown in the penultimate and ultimate examples.

The placement of the accent in HEC languages is often linked to vowel quantity, particularly with the double TVs (§2.2.3). Consequently, even when an accent appears on a single TV, it is historically regarded as a double vowel (Hudson, 1989, p. 6; Hudson, 2006, p. 296). This assumption has also been argued to be true synchronically (Hayward, 1997, p. 113). According to Wedekind (1980, p. 141), long (or double) vowels do not occur in word-final position in Sidaama and Gedeo, unless final accent/stress is analyzed as "length". For Wedekind, the Sidaama examples with unaccented double terminal vowels should be interpreted as equivalent to accented single vowels. Thus, Kawachi's (2007, pp. 63–66) Sidaama examples could be interpreted as having either unaccented double vowels or accented single vowels. However, these alternative interpretations are not strictly necessary since vowel quantity and accentuation are not mutually exclusive, at least from a historical perspective. Additionally, Hudson's general rule (1976, p. 248) states that stress tends to fall on "strong" syllables, providing a broad framework for understanding the weight of stressed syllables in the HEC languages.

However, previous studies (Hudson, 1976, 1989, 2006; Hayward, 1997; Wedekind, 1980; Sasse, 1982) have not taken into account the potential loss of final consonants, leading to exceptions that challenge the rules they establish. In terms of vowel quantity and syllable position, the accent typically falls on the final heavy syllable. In some individual languages, there are exceptional cases where the accent is placed on an open ultimate syllable with a final single vowel (considered a light syllable). These cases may either derive from a historical or underlying presence of double vowels, as previously suggested, or result from apocope, where the final consonant is lost, resulting in an accented open light syllable (see §4.1.4.1, Genitive Vowel and Suprafix and §4.1.4.4).

Although the above examples are confined to the citation forms, the same general rule of accent placement is assumed for the forms extended by the addition of suffixes, at least underlyingly or historically. That does not preclude the possibility that the individual languages have innovated new systems of accent placement while keeping in mind the historical origin of TVs as remnants of fossilized suffixes (§§2.2.3, 4.1.5.2).

The previous discussion shows the correlation between sonority levels and the placement of the accent. In HEC languages, both voiced and voiceless TVs exist. Typically, the voiceless TVs are not accented and may even go untranscribed in some instances, e.g., Kambaata *ísu* ~ *ís* 'he' (nom). The

placement of the accent on the nominative and accusative pronouns in HEC serves as a clear example of how sonority influences accent placement.

	Voiced -V	**Accusative**	**Voiceless -V**	**Nominative**
Burji	**ʔiʃeé**	‘her’ (acc)	**ʔíʃ̥**	‘she’ (nom)
	ʔisí[83]	‘him’ (acc)	**ʔísᵢ̥**	‘he’ (nom)
Hadiyya	**ʔíseᵢ̥**	‘her’ (acc)	**ʔísᵢ̥**	‘she’ (nom)
	ʔít’t’ᵢ̥	‘him’ (acc)	**ʔít’t’ᵒ̥**	‘he’ (nom’
Sidaama	**isé**	‘her’ (acc)	**íse**	‘she’ (nom)
	isó	‘him’ (acc)	**ísi**	‘he’ (nom)
Gedeo[84]	**iseé**	‘her’ (acc)	**íse**	‘she’ (nom)
	isé-(eʔᵉ̥)		**íse**	
	isoó	‘him’ (acc)	**ísi**	‘he’ (nom)
	isó-(oʔᵒ̥)		**ísi**	
Kambaata	**isé-ta**	‘her’ (acc)	**íse**	‘she’ (nom)
	isú	‘him’ (acc)	**ís(u)**	‘he’ (nom)

Some case vowels are accented in HEC. The accusative vowels are found accented in Burji, Sidaama (Kawachi, 2007, pp. 180, 68–69), Gedeo (Eyob Kelemework, 2015, pp. 54, 112; Dawit Tilahun, 2019, p. 137), and Kambaata (Treis, 2008, pp. 108, 79). The genitive case is also associated with accent in Sidaama and Kambaata (§4.1.4). The nominative counterparts of the above forms are not accented because the case vowels are extremely short and devoiced (Wedekind, 1980, p. 137; Tadesse Sibamo, 2015, p. 86). Although the above examples are only the third-person singulars, the same applies to the other persons.

The manifestations of the apocope process are described for various reflexes of TVs as lax, devoiced, and almost inaudible in Burji, Hadiyya, and Kambaata (Hudson, 1976, p. 250), devoiced in Burji, Sidaama, and Gedeo (Wedekind, 1980, p. 137), shortened and devoiced in Kambaata as well as in K’abeena and Alaaba (Treis, 2008, p. 20), more specifically, devoiced, extra-short, barely audible, and tending to be realized as zero in allegro speech in Kambaata (Treis, 2008, p. 20, fn. 6), similarly lost in continuous speech in

[83] Tesfaye Baye’s (2015, p. 140) transcription of *ʔisᵢ̥́* ‘him’ (acc) with devoiced and accented terminal vowel is likely a mistranscription. Hence, it is transcribed without the devoicing symbol here.

[84] The first-listed Gedeo forms in each example are from Dawit Tilahun (2019), and the second forms with the glottal stop are from Eyob Kelemework (2015). The sources sometimes do not indicate the accent on the form but describe it, so I added them to all forms for clarity.

Hadiyya and less commonly in Kambaata (Hudson, 1976, p. 250), and extremely short and devoiced in Hadiyya (Tadesse Sibamo, 2015, pp. 27, 68, 86).

The quality and quantity of the syllable nucleus are crucial for explaining the observed accent patterns in the two designated syllable positions (penultimate and ultimate) across HEC. For instance, double vowels tend to carry the accent more frequently than single vowels, while voiceless vowels do not receive any accent. The following illustration roughly incorporates these three sonority variables alongside the positional variables, which work together to determine accent placement. The focus is on the accent placement on the ultimate position (terminal vowels) and the relative placement on the penultimate syllable. In the following schema, "**x**" indicates that an accent is possible on that syllable type, while "**-**" indicates that an accent is not attested.[85]

	Terminal Vowels	**Ultimate Syllable**	**Penultimate Syllable**
Sidaama	Double	**x**	**-**
	Single voiced	**-**	**-/x**
	Single voiceless	**-**	**x**
Gedeo	Double	**-**	**x**
	Single voiced	**-/x**	**-**
	Single voiceless	**-**	**x**
Burji	Double	**x**	**-**
	Single voiced	**-/x**	**-/x**
	Single voiceless	**-**	**x**
Kambaata	Double	**x**	**-**
	Single voiced	**-/x**	**-/x**
	Single voiceless	**-**	**x**
Hadiyya	Double	**-**	**x**
	Single voiced	**-**	**-/x**
	Single voiceless	**-**	**x**

Read as a broad tendency rather than a categorical rule, the schema suggests a diachronic drift: systems that still permit accented final VV (as in Sidaama) look more conservative, while systems with reduced or devoiced TVs

[85] These might not apply to every instance in the individual HEC languages, but they are proposed as a general historical framework in light of the examples discussed above. Grammars should be consulted for language-specific rules.

increasingly shift prominence to the penultimate syllable. Burji and Kambaata appear intermediate in this respect,[86] and Hadiyya represents the most advanced reduction of TVs, with no accented final vowels in the data considered here. The "**-/x**" under single voiced TV indicates that the syllable can be accented or not. The conditional relationship shows that there must be one accented "**x**" syllable (ultimate or penultimate), so there are no instances of the absence of accent on both the ultimate "**-**" and penultimate "**-**" syllables or of the presence of accent (high prominence) on both the ultimate "**x**" and penultimate "**x**" syllables of the same word.

In cases where there is "**-/x**" for both single voiced and voiceless TVs in Burji and Kambaata, the presence of "**-**" in the voiced TV (i.e., ultimate) means "**+**" (i.e., accented) penultimate and vice versa. Therefore, this schema illustrates a shift in accent placement from the ultimate syllable (TVs) to the penultimate syllable, which clearly occurs in all cases of single voiceless TVs and the cases where TVs are not accented.

PHEC accent placement followed a weight-sensitive pattern, falling on the ultimate heavy syllable (that containing a double vowel or being closed by a consonant) and on the penultimate syllable otherwise. The modern reflexes show systematic variation: Sidaama and Gedeo retain ultimate accent on double TVs; Burji and Kambaata show intermediate patterns with both double and single accented voiced TVs; Hadiyya exhibits the most advanced reduction, with accent consistently on the penultimate syllable due to complete devoicing of TVs. These patterns support the moraic analysis of HEC vowel quantity adopted throughout this study.

2.5 Summary and Conclusions

This section synthesizes the phonological findings presented in §§2.1–2.4 and contextualizes the reconstruction of PHEC within the broader HEC grammatical framework.

The Proto-Highland East Cushitic phonological system includes a five-vowel inventory with contrastive quantity, relatively symmetrical consonant subsystems, and distinctive prosodic patterns. PHEC reconstructions explain

[86] This analysis differs from Hayward's (1997) proposal regarding the reconstruction of both double and single terminal vowels, which he proposes even for PEC. However, as previously mentioned, this proposal does not take into account the possibility of apocope resulting in final consonant loss. Additionally, Hayward's hypothesis imposes the influence of North Omotic before any attempt to provide an internal explanation.

the reflexes across five daughter languages (Burji, Hadiyya, Kambaata, Sidaama, and Gedeo) using comparative evidence from Cushitic cognates, morpheme boundary alternations, and distributional constraints. The consonant inventory shows series in labial, alveolar, post-alveolar, velar, pharyngeal, and glottal positions. Non-pulmonic consonants contrast at three places of articulation: alveolar (**t'**), post-alveolar (**c'**), and velar (**k'**), with obstruent (sibilant and velar) harmony processes further complicating their reflexes. These patterns distinguish inherited features from areal innovations. The study addresses diverse phonological developments by analyzing correspondence patterns, minimal pairs, and alternations at morpheme boundaries.

2.5.1 Consonant System

PHEC had a consonant inventory of 21 phonemes. The obstruents (obst.) and sonorants (son.) constitute two major HEC sound classes.

Table 3: *Proto-Highland East Cushitic Consonants*

		Labial	Alveolar	Post-alveo.	Velar	Phary.	Glottal
Obst.	Stop	***b**	***t, *d**		***k, *g**		***ʔ**
	Glottalized		***t'**	***c'**	***k'**		
	Fricative	***f**	***s, *z**	***ʃ**		***ħ**	***h**
Son.	Lateral & Trill		***l, *r**				
	Nasal	***m**	***n**				
	Glide	***w**		***y**			

This proto-inventory is smaller than the 27 consonants collectively attested across the modern HEC languages (**Table *1***), reflecting the exclusion of contact-induced consonants like **d'** and **s'** (§2.1.2). Voiceless **p** is also not reconstructed as a PHEC phoneme; it appears in Burji primarily through loans from Oromo. Plumonic **c** and **j** are not reconstructed as PHEC phonemes; they appear as derived or loaned sounds in individual languages.

The phonological system reveals both stable correspondences across HEC languages and regular changes in specific languages, as summarized in the following paragraphs.

PHEC had voiceless stops (***t, *k, *ʔ**) and voiced stops (***b, *d, *g**) at three places of articulation. Non-pulmonic ***t'**, ***c'**, and ***k'** are also present in three positions. The labial stop **b** shows variation with **β** and **w** in intervocalic

positions in Burji, Sidaama, and Gedeo. Non-pulmonic ***c'** shows regular correspondence patterns across the HEC languages and has numerous Eastern Cushitic cognates, supporting its status as inherited, though some instances may reflect contact influence. The glottal stop ***ʔ** is often lost post-vocalically, but functions as a suffix in various verbal derivations supporting its phonemic status.[87]

PHEC has three primary fricatives: ***f**, ***s**, ***ʃ**. Unlike ***s**, the evidence for the reconstruction of ***ʃ** is more tenuous; its correspondences are sporadic. Most **ʃ** reflexes suggest a merger with **s**, though some cases preserve the **ʃ**. The voiced fricative ***z** has been maintained only in Kambaata (and Alaaba); otherwise, it merges with **d** in Burji, Hadiyya, Sidaama, and Gedeo.

Similar to the glottal stop ***ʔ**, the glottal fricative ***h** is often lost post-vocalically. The pharyngeal fricative ***ħ** represents an arguable reconstruction based on initial **h**/**w** correspondences and non-initial traces. External evidence from East Cushitic cognates showing **ħ** supports its reconstruction, though its attestation remains limited (§2.1.5.3). The distribution suggests ***ħ** and ***h** merged word-initially before ***u** and ***a**, creating the frequent PHEC ***ħu** and ***ha** sequences, with distinctiveness preserved only in specific environments.[88]

PHEC had two nasals, ***m** and ***n**, with stable correspondences across the daughter languages. PHEC ***n** is rare in initial position except in pronominal forms. The lateral ***l** and trill ***r** yield consistent reflexes in HEC, with specific mutual sound changes in Hadiyya.[89]

The glides ***w** and ***y** show variable reflexes, often being lost intervocalically or word-finally. They also participated in various morphophonological

[87] The following sections discuss stops and non-pulmonics across the three chapters: Phonology §2.1.1.1, §2.1.2, §2.1.5; Boundary alternations §3.1.1.1, §3.1.1.5, §3.1.2.1, §3.3 (stop assimilation), §3.2 (stops lenition); Morphology §§4.2.3.1, 3.2.2 (verbal derivation).

[88] Sections relevant to the fricatives across the three chapters include: Phonology §2.1.1.2, §2.1.3, §2.1.5.2, §2.1.5.3; Boundary alternations §3.2.2 (velar lenition), §§3.1.1.3, 3.1.2.2 (fricative assimilations), §3.3.1.4 (fricative glottalization); Morphology §4.2.2 (causative).

[89] Discussion of nasals and liquids is found in Phonology: §2.1.1.3, §2.1.1.4, §2.1.3; Boundary alternations: §§3.1.1.2, 3.1.1.4 (nasal assimilation), §3.4.2 (nasal insertion), §3.3.2 (nasal metathesis), §3.1.2.3 (nasal palatalization); and Morphology: §4.1.4.1, §4.1.1.2 (noun morphology), §4.2.1 (verb morphology).

processes, e.g., ***y** triggered palatalization of consonants, while ***w** had rounding effects.[90]

2.5.2 Reflexes of Reconstructed Consonants

The reflexes of reconstructed PHEC phonemes show both regular inheritance patterns and instances of contact-induced change. The lists below summarize these reflexes, separating securely inherited phonemes from those with evidence of significant contact influence. The reconstructions follow the evaluation criteria established in §1.5.1, with each proto-phoneme assigned a confidence rating based on the regularity and breadth of its attestation.

The most reliable correspondences occur with stops, nasals, and most vowels. More complex patterns are observed with sibilants, which are affected by both regular sound changes and areal features like obstruent harmony and contact influence. Detailed correspondences and reflexes are presented below:[91]

Table 4: *Reconstructed HEC Consonants and their Correspondences*

PHEC	Burji	PNHEC	Had	Kam	Sid	Ged	Environment	Confidence
***b**	b/w	***b**/***w**	b/w	b/w	b/β/w	b/β/w	#_, V_V	secure
***t**	t	***t**	t	t	t	t	all	secure
***d**	d	***d**	d	d	d	d	all	secure
***z**	d	***z**	d	z	d	d	all	secure
***s**	s	***s**	s	s	s	s	all	secure
***ʃ**	s/ʃ	***ʃ**	s	s/ʃ	s/ʃ	s/ʃ	varies	arguable
***f**	f	***f**	f	f	f	f	all	secure
***l**	l	***l**	l, r	l	l	l	#_, V_V	secure
***r**	r	***r**	l, r	r	r	r	#_, V_V	secure
***m**	m	***m**	m	m	m	m	all	secure
***n**	n	***n**	n	n	n	n	all	secure
***w**	w/ø	***w**/***ø**	w/ø	w/ø	w/ø	w/ø	varies	secure
***y**	y/ø	***y**/***ø**	y/ø	y/ø	y/ø	y/ø	varies	secure

[90] The following are all the glide-related sections in this study: Phonology §2.1.4, §2.1.3, §2.2.2.1; Boundary alternations §3.4.1 (glide insertion), §3.1.2.2 (glide fusion); Morphology §4.1.1.2, 4.1.5.2 (noun morphology), §4.2.4.1 (verb morphology).

[91] The list reflects the general patterns observed; exceptions and dialectal variations may exist, e.g., Kambaata dialectal initial **w** ~ **y**. Consonant loss is sometimes compensated for by epenthetic sounds not listed. Segments and sequences like **c**, **j**, **p'**, **d'**, **ʔm**, **ʔn**, **ʔl**,and **ʔr** result from morphophonemic processes and contact.

PHEC	Burji	PNHEC	Had	Kam	Sid	Ged	Environment	Confidence
***k**	k	***k**	k	k	k	k	all	secure
***g**	g	***g**	g	g	g	g	all	secure
***k'**	k'	***k'/ø**	k'/k/ø	k'/k/g/ø	k'/k/h/ø	k'/h/ø	varies	arguable
***ʔ**	ʔ/ø	***ʔ/*ø**	ʔ/ø	ʔ/ø	ʔ/ø	ʔ/ø	varies	secure
***h**	h/ø	***h/*ø**	h/ø	h/ø	h/ø	h/ø	varies	secure
***ħ**	h, y	***ħ**	w, h	w/y, h	w, ø	w, ø/y	_*u, V_	arguable
***t'**	t'/d'/c'	***t'**	t'	t'	t'/d'/c'	t'/d'/c'	varies	reasonable
***c'**	c'	***c'**	c'	c'	c'	c'	all	reasonable

Reflexes of both ***l** and ***r** in Hadiyya are **l** in initial position and **r** in intervocalic position. The symbol **ø** indicates sound loss, as seen in **h**/**ø**, **w**/**ø**, and **y**/**ø**, which are examples of the loss of the glides in certain contexts, especially in non-initial position. The cases of **b**/**β**/**w**, **ʃ**/**s**, and **k'**/**k**/**g**/**ø** are examples of variation in reflexes depending on various contexts.

The geminates **bb**, **ll**, **rr**, **kk**, and **zz** are found in limited attestation as examples of regular (geminate) consonant sequences. Most, if not all, the clusters result from morpheme boundary alternation formations not part of the root. The same applies to a number of geminates.

The irregular **d'** occurs in Burji, Gedeo, and sometimes Sidaama, with correspondence to **t'** in Hadiyya and Kambaata in certain cognate sets. Many instances of **d'** in Burji, Gedeo, and Sidaama have direct parallels in Oromo. Sidaama shows free variation between **d'** and **t'**, correlating to Oromo influence. Non-initial **d'** often represents a reflexive suffix, which is also found in Oromo.

		Burji	Had	Kam	Sid	Ged	Environment
***t'**	→	**d'**	**t'**	**t'**	**d'/t'**	**d'**	#_
***t'**	→	**c'**	**t'**	**t'**	**c'/t'**	**c'/t'**	varies

The other irregular reflexes of ***t'** appear as **c'**, mainly in Burji and sometimes in Sidaama and Gedeo, with **t'** in Hadiyya and Kambaata. The distribution of these reflexes shows evidence of sibilant harmony in Burji (e.g., [18] *c'eeji* 'blood'); irregular correspondences with **c'** and **t'** in Sidaama and Gedeo; and parallels with areal sibilant patterns seen in **s** ~ **ʃ** variation (see Hayward, 1997, pp. 115–116).

Below is a list of selected common sound changes in HEC.

PHEC → reflex	Environment	Example	Confidence
***l** → Had **r**	V_V	[137] **lal-* → Had *laro* ‘cattle’	secure
***r** → Had **l**	#_	[184] **reɦ ʔ* → Had *leh-* ‘to die’	secure
***rr** → Had **ll**	_[non-initial]	[110] **barr-* → Had *balla* ‘day’	secure
***s** → NHEC **ʃ**	_ʃ	[406] **wees-ʃ-* → Kam *weeʃʃu* ‘ensete tree (Ensete ventricosum)’	secure
***z** → Kam **z**, rest of HEC **d**	all	[118] **zaak-* → Kam *zaaz-* ‘flow’, Ged daak- ‘swim’	secure
***t’** → Burji, Ged, Sid **d’**	sporadic	[250] **t’iib-* → Burji, Ged *d’iib-* ‘push’	tentative
***b** → **β**/**w**	V_V	[3] **aabb-* ‘father’ → Sid *aaβa* ‘father’	reasonable
***b** → **ø**	V_V	[186] PHEC **c’ub-* ‘dip in, immerse’ → Sid *c’uuʔ-* ‘baptise’	secure
***ɦ** → **ø**	V_	[38] PHEC **liɦ-* → Burji *lia*, Sid *lee* ‘six’	secure

Contact influence and cross-syllable obstruent harmony (fn. 36) explain some irregular changes.

Process	Scope	Diagnostics	Example	Confidence
Oromo ***s** → **f**	Sporadic, contact-induced	Irregular correspondences	[20] **sazeɦ-* → Burji *fadiya* ‘three’	secure
Oromo verbal suffix **-at**	Verbal morphology	Correspondence with Oro **-at** forms	[450] Sid *saalfat-* ‘be shy’	secure
d’/**t’** variation	Lexical items	Free variation in the same language	[417] Sid *d’ibba* ~ *t’ibba* ‘disease’	secure
Sibilant harmony	$C_1...C_2$	**t’** → **c’**, **z** → **s** where C_1 is a voiceless obstruent, C_2 is a sibilant	[18] **t’iig-* → Burji *c’eeji* ‘blood’; [20] **sazeɦ-* →	reasonable

Process	Scope	Diagnostics	Example	Confidence
			Kam *sas-* 'three'	
Velar harmony	$C_1...C_2$	**t'** → **k'** where C_1 is a voiceless obstruent, C_2 is a velar.	[18] **t'iig-* → Kam *k'eg-* 'blood', §2.1.2	reasonable

Burji differs from the NHEC languages in several key aspects: (a) unlike NHEC languages, Burji shows evidence of a fossilized **d'-** prefix in verbs like *d'irk'-* (intr) vs. *rik'k'-* (tr) 'return'; (b) Burji consistently shows **d'** where NHEC languages show **t'** in certain cognate sets; (c) Burji shows evidence of Oromo-specific sound changes like ***s** → **f**. These distinctive features highlight the early separation of Burji from the NHEC languages and its constant contact with Oromo.

2.5.3 Vowel System

The Proto-Highland East Cushitic vowel system consisted of five vowels with a phonemic quantity contrast. The quantity contrast is analyzed moraically, with single vowels (V) carrying one mora and double vowels (VV) carrying two morae.

Table 5: *Proto-Highland East Cushitic Vowels*

	Front		Back
High	***i, *ii**		***u, *uu**
Mid	***e, *ee**		***o, *oo**
Low		***a, *aa**	

This vowel system is remarkably stable across the HEC languages, with consistent reflexes in all languages except for some contextually conditioned variations (§2.2). The distinction between single and double vowels is phonemic, as evidenced by numerous minimal pairs in each language. The vowel ***e** is relatively infrequent compared to other vowels.

All HEC languages have terminal vowels (TVs), though their quality and quantity vary (§2.2.3). These TVs can be:

Double vowels (e.g., Sidaama *aroo* 'husband')
Single voiced vowels (e.g., Kambaata *saʔa* 'cow')
Single voiceless vowels (e.g., Burji *ʔiʃi̥* 'she' (nom))

Terminal vowel quality often reflects case markers (nominative, accusative, genitive) or gender markers (§4.1.5). In HEC languages, terminal vowels have undergone various degrees of reduction through apocope, with Hadiyya, in which TVs are consistently voiceless (§§2.2.3, 2.4), showing the most advanced form.

Vowel doubling is a phonological process that often results from the loss of consonants, particularly glottals. For example, in Burji, the preconsonantal stem of *kaʔ-* is *kaa-* 'get up' through compensatory doubling after the loss of the glottal (§3.3.3). This process is observed in other HEC languages in specific alternations at morpheme boundaries.[92]

2.5.4 Phonotactics and Prosody

PHEC syllable structure followed the pattern (C)V(V)(C)(C), with CV and CVC being the most common types. Initial consonant clusters are not permitted, while medial clusters are limited to two consonants and subject to sonority sequencing constraints. Some vowel-initial syllables in PNHEC derive from the loss of **kʼ** and **ħ** (§2.1.5). The HEC languages predominantly feature open syllables, with consonant clusters restricted to specific environments. Word-internal syllables may be closed by sonorant consonants or the first member of double consonants (§2.3.3). Consonant clusters are generally limited to geminate consonants, sonorant + obstruent sequences, and glottal stop + sonorant sequences (**ʔm**, **ʔn**, **ʔl**, **ʔr**). Metathesis patterns distinguish Burji (velar + sibilant → sibilant + velar) from NHEC (sonorant + glottal → glottal + sonorant), supporting the primary HEC subgrouping (§5.3.1). Non-initial position serves as an innovation environment for many consonants, allowing for greater diversity in consonant distribution. Therefore, the relevant phonotactic sections include §§2.3, 2.4 (Phonology), and the whole upcoming alternations at morpheme boundaries chapter (Ch. 3).

In Highland East Cushitic, accent/stress is associated with the most prominent syllable or a higher pitch. Accent placement in the HEC languages is primarily confined to the penultimate and ultimate syllables. While accent predominantly falls on the penultimate syllable synchronically, this is likely due to historical changes such as the reduction of double terminal vowels and

[92] The HEC vowels are discussed in Phonology §2.2; Alternations at Morpheme Boundaries §3.4.1.1 (vowel epenthesis), §3.3.3 (compensatory doubling); and Morphology §4.1.5 (nominal suffixes).

apocope. Historically, the accent falls on the ultimate heavy syllable, which includes syllables with double vowels and closed syllables.

Accent placement interacts with vowel quantity and the phonetic status of the terminal vowel. Double vowels attract as accent more readily than single vowels, voiceless final vowels are unacceptable accent-bearing units, and in Hadiyya, the extensive reduction of final vowels effectively eliminates final accent in the material surveyed here. Taken together, these properties offer a plausible route from a historically final-heavy target to the synchronically widespread penultimate pattern.[93]

The cross-referenced sections show that PHEC alternations at morpheme boundaries (Ch. 3) involves several systematic processes that affect the realization of phonemes in morphological contexts: common assimilation processes, especially involving sibilants and nasals, palatalization before front vowels in Burji and the glide ***y** in the other HEC languages, affecting particularly alveolar consonants, epenthesis of vowels to break up impermissible consonant clusters, and epenthesis of nasals to compensate for degemination. The key morphological elements (Ch. 4) that trigger such morpheme boundary alternations include: the causative ***-s**, which triggers assimilation and palatalization; the reflexive ***-ʔ**, which interacts with stem-final consonants; and the singulative ***-ʃ**, which triggers palatalization.

The shared phonological innovations are documented in this chapter 2 and expanded in the following chapter 3. They constitute key isoglosses that, together with the morphological and lexical evidence presented in Chapter 4, establish Proto-Northern Highland East Cushitic as a genetic subgroup within HEC. The phonological evidence thus provides the foundation for the subgrouping proposed in §5.3.

[93] Details and discussion of HEC accent are found in Phonology §§2.3-2.4; Boundary alternations §3.2.2 (prosodic effects); and Morphology §4.1.4.4 (suprafix).

3 Alternations at Morpheme Boundaries

This chapter examines the phonological changes that occur specifically at morpheme boundaries in Highland East Cushitic. These are phonological processes, but unlike the sound patterns documented in Chapter 2, they are triggered by morphological operations, such as the addition of suffixes, the concatenation of stems and affixes, and the interaction of roots with derivational morphology. Because these phenomena require reference to phonological processes and, in some cases, to morphological contexts.

Transcription symbols and data presentation/referencing conventions follow Chapter 1 (§1.5.3), including the use of in-text or footnoted citations for examples drawn directly from grammars rather than from database entries.

This chapter serves a dual purpose. First, it documents and compares the sandhi processes that have shaped HEC word forms across different stages. These include assimilation, palatalization, fusion, elision, metathesis, and various epenthetic strategies. Some of these processes are inherited from PHEC, as evidenced by their attestation in both Burji and the NHEC languages; others are innovations at the PNHEC level or within individual HEC languages. Second, and perhaps more importantly, the chapter demonstrates how recognition of derivational history expands the set of identifiable cognates. Lexemes that appear irregular or unrelated at first glance often prove to be regular once the morphological history of their reflexes is taken into account. A Sidaama lexeme showing a glottalized consonant, for example, may correspond regularly to a pulmonic consonant in sister languages once the fusion of a stem-final obstruent with the derivation suffix **-ʔ** is recognized.

The chapter is organized by the consonants undergoing morpheme boundary alternations. Section 3.1 treats the pulmonic alveolars (**t**, **n**, **s, ʃ**), documenting assimilation of the pronominal suffixes **-t** and **-n** to preceding consonants, partial assimilation of stem-final **m** to the following alveolars, as well as palatalization patterns in different contexts. Section 3.2 examines the velars (**k**, **g**), covering palatalization (particularly in Burji) and intervocalic lenition. Section 3.3 addresses processes involving the glottal stop, including fusion with preceding obstruents (yielding glottalized consonants), metathesis with preceding sonorants, and hiatus repair. Section 3.4 documents sonorant epenthesis (of, e.g., **-y-**, **-i-**, and other sonorants) used to repair phonotactically illicit sequences. Finally, Section 3.5 synthesizes the evidence, summarizing which processes can be reconstructed to PHEC, and

which represent PNHEC innovations. Accordingly, the discussion tracks a broad division between obstruent-driven alternations (Sections 3.1–3.3) and sonorant-driven repairs (Section 3.4), culminating in a comparative synthesis in Section 3.5.

3.1 Pulmonic Alveolars

The pulmonic alveolars **t**, **n**, **s**, and **ʃ** are frequently involved in morpheme boundary alternations across HEC. These consonants occur in key grammatical morphemes: **-t** and -**n** mark pronominals (2nd/3rd person), **-s** indicates causative derivation, and **-ʃ** functions as a singulative and deverbative. Their alveolar place of articulation makes them susceptible to assimilation, palatalization, and fusion processes at morpheme boundaries. This section documents these alternations by process type, beginning with assimilation (§3.1.1), then palatalization (§3.1.2).

3.1.1 Assimilation

This section examines assimilation processes affecting alveolar consonants at morpheme boundaries. Five main processes are documented: the full assimilation of pronominal **-t** to stem-final consonants (§3.1.1.1), a process restricted to the NHEC languages; the assimilation of plurative nasal **-n** after sonorant-final stems (§3.1.1.2), reconstructable to PHEC; the assimilation to the verbal suffix **-s** (§3.1.1.3) in Burji and Kambaata; the partial assimilation of stem-final nasal **-m** before alveolar suffixes (§3.1.1.4); and the affrication and allomorphic variation of the nominal derivational suffix ***-ʃ** (§3.1.1.5). These processes are critical for understanding surface forms in daughter languages and reconstructing underlying morpheme shapes.

3.1.1.1 Pronominal t

In the NHEC languages, the second and third-person suffix **-t** fully assimilates to the final consonant of the stem.[94]

Had	**d** + **t** → **dd**	*sid-titto* → *sidditto* ‘you (sg) saw’
Kam	**k** + **t** → **kk**	*sok-tenta* → *sokkenta* ‘you (pl) sent’
Sid	**g** + **t** → **gg**	*ag-tú* → *aggú* ‘she drank’
Ged	**s** + **t** → **ss**	*has-tee* → *hassee* ‘she, seeking’

[94] Data source: Hudson (1976).

In Gedeo, only partial voicing assimilation occurs (Hudson, 1976, p. 250). A similar process has been attested in Burji (Tesfaye Baye, 2015, p. 69) as shown in the following examples:

Ged	**n** + **t** → **nd**	*egen-teeni* → *egendeeni* ‘she knows’
Ged	**l** + **t** → **ld**	*ofol-toot’t’e* → *ofoldoot’t’e* ‘don’t sit’
Burji	**n + t → nd**	*k’un-* ‘pound’, *-tooni* (3sg.f.jussive) → *k’undoóṇ̩* ‘let her pound’

The difference between the two processes in Gedeo is that the full assimilation takes place when the stem-final consonant is an obstruent, while the partial assimilation occurs after sonorants.

Gedeo presents another notable pattern with the **d’** variant of the reflexive suffix. Eyob Kelemework (2015, p. 71) notes that Gedeo **l** and **r** undergo total assimilation to **d’** across all persons, numbers, and genders, e.g.:

Ged	**r** + **d’** → **d’d’**	*hir-* ‘buy’ → *hir-d’-* → *hid’d’-*
Ged	**l** + **d’** → **d’d’**	*kul-* ‘tell’ → *kul-d’-* → *kud’d’-*

The total assimilation of pronominal **-t** is reconstructable to PNHEC, as it is systematically attested in all four NHEC languages but absent from Burji.

3.1.1.2 Nominal n

Nasal assimilation is widespread across the HEC languages, affecting the plurative pronominal suffix **-n**. This suffix fully assimilates to the stem-final sonorant (nasal, lateral, or tap). This process applies across all HEC languages with the exception of Kambaata (Hudson, 1976, p. 250).

Burji	**r** + **n** → **rr**	*mar-na* → *marra* ‘we will go’
Had	**r** + **n** → **rr**	*mar-nummo* → *mallummo* (***r** → Had **l**) ‘we went’
Sid	**m** + **n** → **mm**	*um-noomo* → *ummoommo* ‘we have dug’
Ged	**l** + **n** → **ll**	*ofol-neeni* → *ofolleeni* ‘they sat’

There is one instance of full assimilation of the plurative ***-n** to a stem-final obstruent **s**.

PHEC	***s** + ***n**	[318]	**is-n-ʔ-* ‘their’ (3pl.poss)
Had	**s + n → ss**		*ʔissúwwi* ‘their’
Kam	**s + n → ss**		*issoʔóo* ‘their’
Ged	**s + n → ns**		*insaʔnneke*, *insaʔnnetʼtʼe* ‘their ’
Sid	**s + n → ns**		*-nsa* ‘their’
Burji	**s + n → s-i-n**		*ʔisiná-* ‘their’

In this instance, ***sn** becomes Hadiyya and Kambaata **ss**, while it is retained with metathesis as **ns** in Gedeo and Sidaama. Notably, Burji deviates from the two NHEC patterns by inserting a vowel **-i-** in **is-n → ʔisiná-* ‘their’ above (§3.4.1.1), suggesting a different strategy for resolving the consonant cluster.

Again, the difference in sonority between the stem-final and suffix consonant is what distinguishes the two processes across HEC. The stem-final sonorant plus sonorant **-n** yields full assimilation in four HEC languages, while the stem-final obstruent plus the sonorant **-n** results in various processes (assimilation, metathesis, and epenthesis). The regularity of full assimilation of sonorant + pronominal ***-n** is attested in Burji and in Hadiyya, Kambaata, and Sidaama. This distribution beyond the NHEC branch establishes the process as reconstructable to PHEC. The two processes applied to the pronominal suffixes in §3.1.1.1 and §3.1.1.2 can be formalized as: **-C → C_i / C_i+_**, if they belong to the same class of sonority.

3.1.1.3 Verbal s

Verbal derivation **-s** includes the causative (§4.2.2) and reduced variants of the extended denominative (§4.2.5.7), but its surface realizations frequently reflect morphophonological interaction at morpheme boundaries. In several languages (e.g., Kambaata and Sidaama), suffixation of **-s** triggers full regressive assimilation of a preceding stem-final consonant, yielding a geminate **ss**, as illustrated below.[95]

Kam	**b + s → ss**	*beeb-* ‘be covered’ → *bees-s-* ‘cover’
	z + s → ss	*zaaz-* ‘flow’ → *zaas-s-* ‘make flow’
Sid	**b + s → ss**	*bub- ‘to burn’* → *buʃ-ʃ-* ‘cause to burn’
	d + ʃ → ʃʃ	*godoodo* ‘the shout of cattle’ → *godooʃʃ-* ‘to cause to shout’
Burji	**b + s → ss**	*jib-* ‘run’ → *jiss-*
	d + s → ss	*sood-* ‘stay up all *night’* → *sooss-* ‘make/cause stay on duty the whole night’

[95] Data sources: Kambaata (Treis, 2022) and Burji (Tesfaye Baye, 2015), Sidaama (Indrias Xa’miso et al., 2007; Kawachi, 2007). Glossing is only provided when found in the source.

Burji shows metathesis rather than full assimilation when the stem-final obstruent is a velar, e.g., [701] *jig-* 'to demolish/destroy' → *jisk-*. Therefore, metathesis is the logical intermediate stage for all the obstruents + **s** in Burji, i.e., it is retained with the velars but followed by subsequent full assimilation with the labials and alveolars. Burji does not participate in the full assimilation of the obstruents discussed in §3.1.1 either. Therefore, the full assimilation process of the stem-final single obstruent + verbal ***-s** is inherited from PNHEC in Kambaata and Sidaama and innovated independently in Burji in a more restricted environment (§3.1.2.2).

The sonorants **r**/**l** + verbal **s** display diverse patterns of retained **rs**/**ls** and innovated assimilation and epenthesis.[96]

Had	**r** + **s** → **ss**	*mar-* 'to go' → mas-s- 'to take, to let go'
	l + **s** → **ls**	*t'allal-* 'to slip' → *t'allal-s-* 'to let slip'
Kam	**r** + **s** → **ss**	*mar-* 'go' → *mas-s-* 'take (somewhere)'
	l + **ʃ** → **ʃʃ**	*mool-* (intr) 'dry' → *mooʃʃ-* 'to become dry'
Sid	**r** + **s** → **ss**	*mar-* 'go' → *mas-s-* 'to take away'
	l + **ʃ** → **ʃʃ**	*mool-* (intr) 'dry' → *mooʃʃ-* 'to become dry'
Burji	**r + s** → **rs**	*mar-* 'to go' → *mar-s-* 'make (someone) go'
	l + **s** → **ls**	*ilal-* 'to open eyes' → *ilaal-s-* 'to display'
Ged	**r + s** → **ris**	*mar-* 'go' → *mar-is-* 'cause to go, accompany'
	l + **ʃ** → **lc**	*mal-* 'find a way out' → *mal-c*

In the above examples, Hadiyya, Kambaata, and Sidaama assimilate the **rs** sequence fully into **ss**. However, such a full assimilation of the sonorant **r** and **s** does not apply to Burji and Gedeo, because Burji retains the sequence and Gedeo inserts **-i-** between the segments. For the sequence **ls**, Burji and Hadiyya retained the sequence. There is palatalization of **s** to **ʃ** (§3.1.2.2) in the other three languages, but only Sidaama and Kambaata show full assimilation of **l+ʃ** into **ʃʃ**. Gedeo affricated the **ʃ** into **c** without full assimilation. Unlike the obstruents + **s** above, the reflexes of the sonorants + **s** present various manifestations; therefore, they are classified as independent developments in each daughter language. This conclusion also complies with the allomorphs of the verbal **-s** and their environments established in §3.1.2.2.

[96] Data sources: Hadiyya (Ritter, 2007), Kambaata (Treis, 2022), Sidaama (Indrias Xa'miso et al., 2007), Burji (Getahun Addo, 2017, pp. 649, 651), Gedeo (Eyob Kelemework, 2015, p. 214).

3.1.1.4 Stem-final m

Several alveolar-initial suffixes have contributed to the partial assimilation of the stem-final ***m** to **n** in HEC.

***m + *t**, ***s**, ***c'**, ***ʃ → n**

The following examples show the reconstructed words with stem-final ***m** in the first column, retained **m** in the second column, and assimilated **n** in the third column:

[198]	m	→ n	Gloss	[142]	*m	→ n	Gloss
PHEC	***tum-am-ʃ-, *tum-t-, *tum-t-i-ʃʃ-**	**-**	'blacksmith'		***tam-n-**	**-**	'ten'
Burji	**-**	**tuntoo**	'blacksmith'			**tanna**	'ten'
PNHEC	***tum-am-ʃ-, *tum-t-, *tum-t-i-ʃʃ-**		'blacksmith'		***tom-n-**	**-**	'ten'
Had	**tumaanco**		'blacksmith'		**tommo**	**-**	'ten'
Kam	**tumaan-cu**		'blacksmith'		**-**	**tonn-a**	'ten'
Sid	**-**	**tunto**	'blacksmith'		**-**	**tonne**	'ten'
Ged	**-**	**tunto**	'blacksmith'		**tomme**	**-**	'ten'

Unassimilated **m** is also found in Oromo *tumtuu* 'smith'. In [142], only Sidaama and probably Kambaata have an instance of assimilated **m** before **n**, whereas Hadiyya and Gedeo assimilated **n** to the preceding **m**, stimulated by the matching sonority. Kambaata *tonn-a* is found only in Sasse (1982, p. 176), while other sources have *torduma* 'ten'. According to Treis (p.c.), the correct transcription is *too-na,* and the translation should be 'ten + X', as it is only used in '11, ... 19', but not in '10'.

[675]	*m	→ n	Gloss	[170]	*m	→ n	Gloss
PHEC	***k'oom-**	**-**	'calabash'		***sum ʔ**		'urinate'
Burji	**k'oomoo**	**-**	'calabash'		**-**	**sinʔ-**	'urinate'
PNHEC	***k'oom-**	**-**	'calabash'		***ʃum ʔ**		'urinate'
Had	**k'ooma**	**-**	'calabash'		**ʃum-eʔ-**		'urinate'
Kam	**k'oomu**	**-**	'calabash'		**ʃum-aʔ-**		'urinate'
Sid	**-**	**k'oon-co**	'calabash'		**ʃumi**	**finc'i**	'urinate'
Ged	**k'omma**	**k'oon-jo**	'calabash'			**ʃiʔn-**	'urinate'

In [170], Oromo *finc'-eessa* 'urinate' is likely the source of the second Sidaama variant *finc'i*. Burji and Gedeo **n** is likely due to analogy with the Oromo form, which also explains the irregular vowel reflex **i**.

[363]	***m**	**→ n**	**Gloss**	**[607]**	**m**	**→ n**	**Gloss**
PHEC	***om-t-**	**-**	'five'		***ħuy ʔ-s-**		'to fill'
Burji	**umutta**	**-**	'five'		**-**	**(huuss)**	'fill'
PNHEC	**-**	***ont-**	'five'		***ħuym-s-**	**-**	'to fill'
Had	**-**	**onto**	'five'		**-**	**wanʃ-**	'fill'
Kam	**-**	**ontu**	'five'		**-**	**wiinʃ**	'fill'
Sid	**-**	**onte**	'five'		**woʔm-is-**	**wonʃ-**	'fill'
Ged	**-**	**onde**	'five'			**wonʃ-**	'fill'

The partial assimilation ***m** → **n** is conditioned by alveolar-initial morphemes: ***-t** ([675], [198], possibly [363]), **-ʃ** ([607]), **-n** ([142]), and **-c'** ([170]). For example, in Burji, **mt** became **nc** in *gama-* 'bite' → *gánc̣i*, where the **c** is from the consonant **-t** of the subject agreement marker, which palatalizes in such a post-consonantal context (Tesfaye Baye, 2015, p. 72) before **-i** (§3.1.1.5). Kawachi (2007, pp. 49–50) documents a synchronic assimilation of stem-final **m** to **n** before **t** in Sidaama. Although it is not applied to the reconstructed PHEC above, such a widespread partial assimilation process also dates back to PHEC.

3.1.1.5 Nominal ʃ

Hudson (1989, pp. 11–12) reconstructs PHEC *c [tʃ], describing its variants as "c is [č:] between vowels and (single) [č] elsewhere, except in Burji and Gedeo, which seem to have intervocalically a marginal contrast of length between c and cc". However, the affricate **c** shows a defective distribution across HEC languages, including Burji (Tesfaye Baye, 2015, p. 32), Sidaama (Anbessa Teferra, 2000, p. 13), Kambaata (Treis, 2008, p. 25), and other HEC languages (Wedekind, 1990, p. 51). The affricate is restricted to the context of derivational morphemes (singulative §4.1.1.1 and deverbative §4.1.2). It has two more allomorphs, **ʃ** and **j**, and only the post-alveolar **ʃ** is a phoneme. Consequently, the affricates (**c** and **j**, representing [**tʃ**] and [**dʒ**])) are not reconstructed as independent PHEC phonemes; rather, they are analyzed as recent developments from ***ʃ**.

The shape and attestational evidence of the singulative (§4.1.1.1–4.1.2) overlaps substantially with that of the deverbative suffix discussed in §4.1.2. Both suffixes exhibit the same reflexes, and in many lexical items, it remains

unclear whether such a suffix has a singular or deverbalizing function. This overlap suggests a possible shared origin or early merger of the two morphemes. Therefore, they are discussed together in this section.

The following sections discuss two hypotheses for the origin of the (singulative/deverbative) derivation suffixes: The first hypothesis reconstructs an original ***t** by proposing an expressive palatalization to explain the affricate **c** allomorph, while the second hypothesis argues for an original post-alveolar ***ʃ** affricated into **c** and assigns multiple sources for the instances of assumed fossilized ***t**. The second hypothesis is more plausible, as demonstrated by the discussion and supporting evidence.

Affricated c

The first hypothesis relies on comparative evidence, such as [215] Gedeo *manj-icco* ‘woman’ with Hadiyya *meent-ícco̥* and Kambaata *meent-iccu-ta* ‘woman’, which suggests that a PHEC derivational suffix, in this case the singulative suffix ***-t** underwent palatalization to ***-c** rather than affrication from an older ***ʃ**. Palatalization of the singulative has an expressive purpose (e.g., when the singulative is used as a diminutive in Kambaata (Treis, 2008, p. 141)).[97] According to the first hypothesis, the singulative morpheme ***-t** has two expressive allomorphs ***-i-cc** and ***-c**; their selection depends on the stem-final sound.

Context	Morpheme	Allomorph
everywhere	***-t** →	
after obstruents/clusters		***-i-cc**
after sonorant		***-c**

After stem-final sonorants, the allomorph of the singulative suffix attached is ***-c**; elsewhere, it is ***-i-cc** (Anbessa Teferra, 2000, p. 27; Tadesse Sibamo, 2015, p. 48; Treis, 2008, p. 132), which might have been the case since early stages. The geminated allomorph **-cc** appears preceded by an epenthetic vowel **-i-** (see Hudson, 2006, p. 296; Hudson, 1989, pp. 11–12; Anbessa, 2000, p. 13), and both **-cc** and **-c** are always followed by the terminal vowel **-**

[97] A hypothesis sound symbolic or expressive gemination might also account for discrepancies like the irregular gemination correspondences in Gedeo [367] *mic’c’iir-* ‘twist’, *mic’c’-* ‘squeeze’ vs. Burji *mic’ir-* ‘wring out’ and Burji *hat’t’iʃee-d’-* (cf. *ha**tt**iʃeed’-*, *ha**tt**iʃeys-* (Roba Dame & Wedekind, 2008, p. 37)), Hadiyya *hat’t’eʃ-* vs. Burji *hat’iʃ-*, Hadiyya *hat’iʃʃuy-*, Kambaata *hant’iʃa?*, Sidaama *hant’iʃʃ-ir’-*, *hant’iʃʃo* ‘to sneeze’.

o in the singulative. Further palatalization patterns (**ʃ**, **j**) still prevail when the non-geminate allomorph **-c** meets other stem-final consonants.

Evidence of the proposed derivation of ***t** could be observed in instances of fossilized ***t** corresponding to palatalized **c** or unmarked **ø**.

[215]	**Unmarked**	**-t**	**-c/-cc**	**Gloss**
PHEC	**-**	***meen-t-**	**-**	‘woman, girl’
Burji	**meena**	**mandeyi**	**-**	‘people’, ‘man, person’
Had	**-**	**meent-íccọ**	**meent-íccọ**	‘woman’
Kam	**-**	**meent-iccu-ta**	**meent-iccu-ta**	‘woman’
Sid	**-**	**-**	**manco**	‘man, woman’
Ged	**-**	**-**	**manj-icco**	‘woman’

Example [215] includes the fossilized ***t** and later derived forms.

[175]	**-t**	**-c/-cc**	**Gloss**
PHEC	***kalt-**	**-**	‘axe, hoe’
Burji	**kaltee**		‘axe’
Had		**kalceʔe**	‘certain part of the plow’
Kam	**kalta**		‘hoe’

Only Hadiyya has a palatalized **c** in the above example [175].

The cognate relationship of the lexemes with **t** is less certain in the following example:

[121]	**Unmarked**	**-t**	**Gloss**
PHEC	***gol-**		‘wall’
Burji	**gola**	**godoo**	‘supporting wall’, ‘wall’
Had	**-**	**gorteʔe, gole-ta**	‘wall’
Kam	**gola**		‘horse stable’
Sid	**golo**	**gaata**	‘firewood-place, fence’, ‘wall’
Ged	**golo**		‘light wall/curtain’

The weaknesses of the first hypothesis include (a) non-phonetic conditions (expressive palatalization); (b) the limited fossilized evidence, and a possible alternative explanation for the limited examples; (c) the reconstruction of proto-forms with the affricate **c**, introducing a defectively distributed and isolated sound, and a new isolated PHEC/PNHEC consonant class (§2.5.1); and (d) the fact that the two environments (stem-final after sonorants (***-c**) and elsewhere (***-i-cc**)) leave no room to accommodate a third widely attested variant **ʃ**, with apparent overlap in distribution.

Post-alveolar ʃ

The second hypothesis posits ***-ʃ*** as the proto-nominal derivation morpheme. Despite its limited attestation compared to **-c**, ***-ʃ*** has reflexes in all HEC languages. Therefore, it must have merged with the innovated affricate **-c**.

Singulative **-ʃʃo** is found in some nouns in Sidaama. Anbessa Teferra (2000, p. 27) argues that it is a product of a process by which the stem-final consonant **s** or **d** plus **-co** became **-ʃʃo**. Likewise, in Kambaata, Treis (2008, p. 133) describes the singulative **ʃʃ** as a result of the stem-final single consonant merging with the initial consonant **c** (**ʃ** in this study) of the singulative suffix, hence sometimes the result is full assimilation into **cc**, e.g., **b** + **c** → **cc** in *ibiib-i-ta* → *ibiicc-u-ta* ‘louse’.

Had	**b + ʃ → ʃʃ**	*bagado* → *bagaʃʃo* ‘spear, lance’
Kam	**b + ʃ → ʃʃ**	*zagib-a* → *zagiʃʃ-u* ‘cedar tree’
Sid	**d + ʃ → ʃʃ**	*galad-o* → *galaʃʃ-o* ‘monkey’
Ged	**r + ʃ → rʃ, cc**	*moor-* ‘steal’ → *mod’arʃ-o*, *mod’acc-o* ‘the thief’
Burji	**r + ʃ → rʃ**	*d’oor-oó* → *d’oor-ʃoó* ‘fig tree’

Gedeo **-c** in the above example serves a deverbalizing function, not a singulative as Eyob Kelemework (2015, p. 103) claims.

Burji has very few nouns where the singulative is marked by **-ʃaa** or **-ʃoo** (Tesfaye Baye, 2015, p. 95). In Gedeo, the allomorph **-ʃo** is attached after stems ending in the trill **r** (Eyob Kelemework, 2015, p. 103). Apparently, the same environment, i.e., stem-final **-r**, is where the Gedeo and Burji allomorph **-ʃo(o)** occurs in the above examples. Hadiyya is the only language that has not been reported to have the post-alveolar **ʃ** variant of the singulative. However, the available data, such as the example listed above, and the deverbative examples in Tadesse Sibamo (2015, pp. 40–41) (see §3.1.2.1), prove that there are at least traces of the post-alveolar **ʃ**.

The PHEC nominal derivational suffix ***-ʃ*** has various allomorphs: ***-ʃ, *ʃ-ʃ***, ***-i-ʃʃ***, and ***-c***, highlighting the dependence of the surface form on phonological processes. Therefore, the conditioning environment where each PHEC allomorph occurs provides solid ground for the interpretation of the various reflexes (see §§4.1.1.1, 4.1.2 for details on each suffix):

Stem-final	PHEC	PHEC allomorph	Process
$C_{[+son]}$	**+*ʃ**	→ ***-ʃ**	retention
$C_{[+fricative]}$	**+*ʃ**	→ ***ʃ-ʃ**	fusion/assimilation
$C_{[stop]}$	**+*ʃ**	→ ***-c**	affrication
CC, cC	**+*ʃ**	→ ***-i-ʃʃ**	epenthesis & gemination

The above allomorphs are found in the following contexts: ***-ʃ** is retained after sonorants, becomes ***ʃ-ʃ** after fricatives via fusion or assimilation, becomes ***-i-ʃʃ** with epenthetic **-i-** after geminates (CC)/clusters (cC),[98] or becomes affricated into ***-c** after stops. Synchronically, there is extensive merger of ***-ʃ** with the affricated variant **-c**, driven by the unstable history of post-alveolar ***ʃ** (§2.1.1.2) and generalization pressure of the affrication variant **c**, which was restricted to the environment after stem-final stops. Within the established methodological contact/inheritance continuum (§1.5), areal pressure has contributed further to the generalization of the affricate **-c** (found in Oromo (Andrzejewski, 1960, p. 63; Stroomer, 1987, p. 83)) and even **-t** (found in Oromo (Andrzejewski, 1960, p. 63; Stroomer, 1987, p. 83), Konso (Zaborski, 1986, p. 145; Ongaye Oda, 2013, p. 87), and Diraytata (Zaborski, 1986, pp. 159–160)).

The comparative evidence supporting the reconstruction of nominal derivational ***-ʃ** includes the following deverbative [329] and singulative [190], [65] sets:

	[329]		[190]		[65]	
Burji	**reesi**	‘corpse’	**-**	-	**-**	-
Had	**leeʃa, leeʃʃa**	‘corpse’	**ellin-co**	‘sun’	**faraʃʃo**	‘horse’
Kam	**re(e)ʃa**	‘corpse’	**arr-iccu-ta**	‘sun’	**faaʃʃu**	‘horse’
Sid	**reeʃʃa, reeʃa**	‘corpse’	**arriʃʃo**	‘sun’	**faraʃʃo**	‘horse’
Ged	**reenʃa, renʃa**	‘corpse’	**ariʃʃo**	‘sun’	**faracco**	‘horse’

The easiest explanation for the Burji reflex **s** is that it continues an earlier post-alveolar **ʃ**, rather than deriving directly from the affricate **c** (e.g., [329]). Across HEC, the comparative evidence points to a post-alveolar base: NHEC languages retain **ʃ** as such in the relevant sets, whereas Burji reflects this segment as **s** (see §4.1.2). The affricated allomorph **-c**/**-cc** is seen in languages such as Gedeo (e.g., [65]) and Hadiyya/Kambaata (e.g., [190]).

[98] Some synchronic geminates (**ʃʃ**/**cc**) after the epenthetic **-i-** or **-V** might be (a) fusion with a preceding consonant in concatenated suffixes, (b) a phonotactic reason to create a coda for the preceding syllable and an onset for the terminal vowel, (c) lack of quantity contrast between **c** and **cc** (Hudson, 1989, pp. 11–12), or (d) sound symbolic (expressive), especially in singulative.

All the HEC languages retained **ʃ** in the three cognate sets, except for Gedeo in [65] and Hadiyya and Kambaata in [190], which show the affricated **-c** and **-cc**.

From the perspective adopted here, the alleged fossilized derivational morpheme ***t** (central to the first hypothesis) remains difficult to establish. The clearest case is [215], but the additional support ([175], [121]) is limited, and comparable **t** segments also occur in some deverbative formations (see §4.1.2). Given this distribution, ***t** is best treated as ambiguous at present: depending on the lexeme, it may reflect stem material (e.g., [121]), a distinct derivational element (e.g., [215]), or an assimilatory outcome within the same suffix complex (e.g., alveolar-triggered restructuring of ***ʃ**, as in [307]),[99] or contact influence. Crucially, the presence of occasional **t** does not require the reconstruction of an affricate-only proto-suffix.

The second hypothesis is preferable for three reasons. First, it reconstructs a post-alveolar ***ʃ** at the PHEC level, which is more parsimonious than positing an inherently restricted proto-affricate ***-c**. Second, it derives the widespread **-c** outcomes from independently motivated conditioning (after stops) and from later analogical/areal generalization, rather than treating **-c** as primary. Third, it can incorporate the core observations behind the first hypothesis without contradiction: where **t** is involved, it can be analyzed as an input to palatalization/affrication pathways (cf. §3.1.2.1) that ultimately converge on the same suffix complex.

Affricated j

The voiced affricate **dʒ** (represented as **j**) is infrequent in HEC and is often associated with loanwords. In Gedeo, however, **j** also appears morphologically as **-jo**, which corresponds to the singulative allomorph **-co** found elsewhere in HEC specifically after **n**. Synchronically, this is best understood as voicing of the affricated allomorph (i.e., **n** + **co** → **n** + **jo**), rather than evidence for an independent suffix. The same voiced outcome is also

[99] In PHEC [307] **mit- 'one'* → *Burji micca*; PNHE **mat-*, **mit-* → Had *mato*, Kam *matu*, Sid *mitte*, motto, Ged *mitte*, the derivation suffix is transparent in Burji, but the geminate **tt** in Gedeo and Sidaama could be due to full assimilation to the stem final **t**. Hadiyya and Kambaata cognates show no traces of the derivation suffix and also exhibit a different corresponding vowel, as expected with the numerals.

compatible with the Gedeo deverbative pattern -**an**-**jo** in §4.1.2. Consider the following Gedeo examples:[100]

	n+c	**Gloss**
[675]	**k'oon-jo**	'gourd, vessel'
[602]	**kin-jo**	'stone, rock'
[138]	**man-jo**	'man, person, woman'
[59]	**agen-jo**	'month, moon'[101]

The correspondences in other HEC languages exhibit the affricated **-c**:

Sid [675] *k'oon-co* 'calabash, skull'
Sid [602] *kin-co*, Kam *kin-cu*, Had *kin-c* 'stone, rock'
Sid [138] *man-co*, Kam *man-cu* 'man', Had *man-co* 'man, person'
Had [59] *agan-co*, Burji *agun-coo* 'month'

The affricated singulative allomorph **c** is voiced to **j** in Gedeo in all the above examples.

Compare the deverbalized stems with a suffix **-c** and their plural counterparts in [198].

[198]	**Verbalized**	**Plural**	**Gloss**
Burji	**tunt-icca**	**tunto**	'smith', 'clan of smiths'
Had	**tumaan-co**	**tumaano**	'blacksmith'
Kam	**tumaan-cu**	**tumaannu**	'blacksmith'
Sid	**tumaan-co, tunt-icca**	**tunto**	'blacksmith'
Ged	**tunt-icca**	**tunto**	'blacksmith'

This comparative set is derived from [48] **tum-* 'to pound', so it includes deverbative affricated **-c**, which is formally identical to the singulative allomorph **-c**/**-cc**. However, the two functions are not identical in distribution: in deverbatives, allomorph choice is often more transparently conditioned by preceding concatenated suffixes (see §4.1.2).

[100] Gedeo loans from Oromo show initial and non-initial **j**: [38] *jaane* 'six' ← Oromo *jaʔa*, [155] *jaal-at* 'love, like' ← Oromo *jaall-ad'd'a*, [15] *urje* 'star(s)' ← Oromo *urjii*, [265] Gedeo, Burji, Kambaata *huj-* 'work' ← Oromo *hojii*. Another loan in HEC is Burji *jinjibili*, Had *jinjibila*, Kam *jaanjibeelu*, Sid *janjibeelo* (Treis, p.c.), Ged *gaanjibelo* ← Amharic *zənjəbel* (Leslau, 1976, p. 357); Oromo *jinjibila* 'ginger' (Hudson, 1989, p. 70).
[101] Example [59] *agen-jo* is noted as an Omotic loanword (Sasse, 1982, p. 24), found in Zala, Gofa, Gamu, and Dac'e as *agina* 'moon, month' (Lamberti & Sottile, 1997, p. 281). Additionally, Gedeo [173] *billaanca* 'butterfly' is an exception, having **nc** instead of **nj**, so it might be a recent loan.

3.1.2 Palatalization

Palatalization in the HEC languages affects alveolars in both stems and suffixes under specific structural conditions. The process targets **t** and **s** primarily, and more marginally **n**, yielding palatalized reflexes that are central to the alternations surveyed in §3.1. Because the diachronic history of alveolar fricatives is not transparent in all environments (see §2.1.1.2), palatalized outcomes involving **ʃ** should be interpreted together with other morphophonological processes. This is particularly relevant for causatives, where a palatalized causative **-ʃ** is analyzed in Chapter 4 as an innovation of Kambaata, Sidaama, and Gedeo.

3.1.2.1 Stem-final and Pronominal t

Burji exhibits palatalized reflexes of ***t** as the affricated **c** and the fricative **ʃ** in certain environments. The distinctive palatalization **t** → **c** pattern occurs before the high front vowel **i** (or an older Nominative ***y** §4.1.5.2). Additionally, there are traces of what Sasse (1982, p. 18) describes as an older rule where **t** → **ʃ** intervocalically before **i**.

t + i → ci
Vt + i → ʃi

The palatalization of **t** → **c** occurs in pronominals followed by the feminine nominative suffix **-i**.

	PHEC	t+i	Nominative	t+a	Accusative	Gloss
[135]	***ʔy**	**ʔícc̣**	(1sg.poss.f)	**ʔítt̪**	(1sg.poss.f)	‘my’
[188]	***ni-ʔ-**	**ninci**	(1pl.poss.f)	**ninta**	(3pl.poss.f)	‘our’
[318]	***isn-ʔ-**	**ʔisinácc̣i**	(3pl.poss.f)	**ʔisinátt̪a**	(3pl.poss.f)	‘their’
[366]	***ise-ʔ-**	**ʔiʃeécc̣i**	(3sg.poss.f)	**ʔiʃeétt̪a**	(3sg.poss.f)	‘her’
[281]	***t-**	**cin(d)’i-ʃi**	(pl.f.dem.pron)	**-**	-	‘these’
[418]	***ti**	**ci**	(sg.f.det)	**ta**	(sg.f.det)	‘this’

Intervocalic palatalization of ***t** to **ʃ** is attested in a limited set of pronominals, especially before the high front vowel **i**.

	PHEC	Vt+i	Nominative	Gloss
[27]	***at-i**	**ʔáʃi**	(2sg.nom)	‘you (sg)’
[104]	***ati-n-**	**ʔaʃínu̥**	(2pl.nom)	‘you (pl)’
[418]	***ti**	**-ʃi**	(def.f)	‘the’

Burji **-ʃi** (def.f) (Getahun Addo, 2017, p. 788) is likely from the same stem as **ci** (sg.f.det) above, but has **ʃ** in the suffix before a vowel, e.g., *ama-ʃi* 'the woman'.

Both Hadiyya and Burji show palatalization adjacent to obstruents. Tadesse Sibamo (2015, pp. 40–41, 76) describes a process of unconditional palatalization and gemination in Hadiyya where stem-final alveolars **t**, **d**, and **s** palatalize and geminate to **cc**, **jj**, and **ʃʃ**, respectively. While no added suffix is explicitly mentioned, the morphological process clearly involves nominal formation from verb stems, e.g.:[102]

Stem	**Gloss**	**$C_{[-son]}$+ʃ**	**Gloss**
ʔit-	'eat'	**ʔicca**	'meal'
ʔit-is-	'to feed'	**ʔitíʃʃa**	'causing to eat'
ʔed	'add'	**ʔejja**	'addition'
fiink'ees-	'whistle'	**fiink'eéʃʃa**	'whistling'

The gemination and palatalization suggest that the stem-final **-s** fused with the following nominal derivation (deverbative) suffix ***-ʃ** (§3.1.1.5), yielding geminated **cc**, **ʃʃ,** or **jj**, depending on the quality of the stem-final consonant. Unlike the nominal causative *ʔitíʃʃa* 'causing to eat', in the nominal, there is direct attachment of the suffix **-ʃ** to the root final **-t** in *ʔicca* 'meal' above and Kambaata *icca-ta* 'food' (compare with *it-* 'eat, rip off' → *it-is-* (caus1) 'feed, offer (food)' → *it-siis-* (caus2) 'cause to be ripped off' (Treis, 2008, pp. 9, 432)). Therefore, **-is+ʃ** yielded geminated **ʃʃ**, and **t+ʃ** fused into an affricated **c**.

Burji displays unique patterns in the palatalization and subsequent assimilation of the subject agreement suffixes **-t-** (2sg/3sg.f) and **-t-ng-** (2pl) when attached to a stem ending in obstruents.

tt	→ **cc**, **ʃʃ**	*k'ot-* 'dig' → *k'óccị*	*ʔit-* 'eat' → *ʔíʃʃị*
d't	→ **cc**, **ʃʃ**	*had'ad-* 'throw' → *had'áccị*	*god'-* 'weed' *goʃʃị*
ʔt	→ **ʃʃ**	*ʃoʔ-* 'select' → *ʃóʃʃị*	-
bt, **dt**, **gt**	→ **jj**	*hab-* 'forget' → *hájjị, hajjíngị*	-

While Tesfaye Baye (2015, p. 70, fn. 16) follows Wedekind (1985, p. 127) in describing changes like **tt** to **ʃʃ** as dissimilation, their analysis fails to account for other patterns, such as **ʔt** to **ʃʃ**. The conditioning factors for **cc** and **ʃʃ** are puzzling. The front vowel **-i** should theoretically produce **c**, not **ʃ**, in such a context as discussed above in this section. Therefore, contra Tesfaye Baye (2015, p. 70), one tentative scenario proposes a regular palatalization before

[102] Data source: Tadesse Sibamo (2015, p. 40).

the high front vowel ***-t-i** → ***-c-i**.[103] The subsequent changes that the palatalized **-c** undergoes are conditioned by different stem-final consonants and could be summarized as follows: **c** following sonorants ***C$_{[+son]}$ci**, otherwise yielded an impermissible sequence ***C$_{[-son]}$ci** resolved as (a) fused voiceless geminate **cc** following pulmonic voiceless obstruents (except **ʔ**), (b) fused voiced geminate **jj** following voiced stops, (c) fused glottalized geminate **c'c'** following glottalized consonants, (d) free variants **cc** and **ʃʃ** following voiceless alveolar stops (including **d'**), and (e) geminate **ʃʃ** following the glottal stop **ʔ**.

Tesfaye Baye (2015, pp. 278–283) analyzes the vowel **-i̱** in Burji as a converb marker, as in the following examples of the verb 'say'.

ʔíy-i̱	*ʔíy-ø-i̱*	*ʔíʃ-i̱*	*ʔíy-í-ng-i̱*
say-i̱	say-3sg.m-i̱	say-3sg.f-i̱	say-i-3pl-i̱

The palatalized single **ʃ** of the third-person feminine **-t** is the regular reflex in this intervocalic environment, as discussed above. In the third-person plural, the **-i-** is epenthetic (§3.4.1).

3.1.2.2 Stem-final and Verbal s

The verbal derivational suffix ***-s** has several reflexes in the HEC languages, including a palatalized **-ʃ** in Kambaata, Sidaama, and Gedeo and an affricated **-c** in Gedeo. This subsection outlines the morpheme boundary alternation processes that can yield these reflexes, with particular attention to palatalized outcomes and to stem-final **s** palatalization through fusion with the singulative suffix ***-ʃ**.

Like the nominal derivational ***-ʃ**, the causative (§§4.2.2, 4.1.2) and the reduced variants of the causative denominative (§4.2.5.7) ***-s** markers show overlap in form and distribution. In some lexical items, the two markers are formally identical, which makes it difficult to distinguish causative from grammaticalized denominative derivation on phonological grounds alone. This overlap suggests generalization of the denominative, even though the two were originally distinct; for this reason, I discuss them together here as verbal derivation morphemes.

---

[103] Supported further by lack of palatalization before vowels other than the high front vowel, such as **-tooni** (3sg.f.jussive), e.g., *k'ot-* 'dig' → *k'ottoóni̱*, *duk'-* 'mix' → *dut't'oóni̱*. Both show fusion and gemination with the stem-final consonant only.

Alveolar s

The PHEC verbal derivational suffix ***-s** has allomorphs: ***-s** and ***-**is in addition to PNHEC ***s-s**. Therefore, the conditioning environment of each PHEC allomorph provides insights into the current diverse reflexes and environments (see §§4.2.2, 4.2.5.7 for further details):

Stem-final	PHEC	PHEC allomorph	PNHEC allomorph	Process
$C_{[+son]}$	**+*s**	→ ***-s**	→ ***-s**	retention
$C_{[-son]}$	**+*s**	→ ***-s**	→ ***s-s**	fusion & gemin.
CC, cC	**+*s**	→ ***-i-s**	→ ***-i-s**	epenthesis

The above allomorphs are found in the following contexts: ***-s** is retained after sonorants and receives an epenthetic **-i-** after geminates (CC)/clusters (cC) to surface as ***-i-s**. It becomes ***s-s** after obstruents via fusion and gemination in PNHEC and occurs in Burji with the labials and alveolars only (§3.1.1.3).

The PHEC pattern of **-s** suffixation directly to a stem-final singleton is preserved at least in Burji, Kambaata, and Sidaama, as seen in §3.1.1.3. For remnants of evidence in the rest of HEC, see the reflexes of the PHEC verb [425] **ʔ* 'to do' and its derived causative **ass-* (caus) 'to do' (← **ʔ-s*) and PHEC [388] **y* 'to say, do, be' and its derived causative **is-* (← **-y-s*). The epenthetic **-i-** occurs between the stem-final geminates and clusters, as in the following examples:[104]

Burji	**gg + s → ggis**	*ugga* 'to make somebody haste' → *uggisa* 'to make somebody make someone to haste'
Had	**ll + s → llis**	*dill-* 'to descend' → *dill-is-* 'to move something down'
Kam	**nʃ + s → nʃis**	*ganʃ-* 'cough' → *ganʃ-is-* 'make cough'
Sid	**ʔm + s → ʔmis**	*t'uʔm-* 'to become beautiful/handsome' → *t'uʔm-i-s-*
Ged	**rr + s → rris**	*dirr-* 'go down, descend' → *dirr-i-s-* 'unload, put down'

Although it is more prominent in Hadiyya and Gedeo, there is frequent ***-is** across all environments, particularly after obstruents, even in Burji, Kambaata, and Sidaama. Therefore, the verbal allomorph **-is** is the most widespread and attested in all the HEC languages, including Gedeo, which has no retained allomorph **-s**. The source of instances of **-is** in environments other than those specified above is attributed to the identical form and function of the causative ***-i-s** and the derived PHEC verb **is-* meaning 'to make' (from the verb **y-* 'to

[104] Data sources: Burji (Getahun Addo, 2017, p. 835), Hadiyya (Tadesse Sibamo, 2015, p. 127), Kambaata (Treis, 2022, p. 9), Sidaama (Kawachi, 2007, p. 55), Gedeo (Eyob Kelemework, 2015, p. 183).

say, do, be'). Therefore, not every **i** preceding the allomorph **-s** is epenthetic, because **i** has also historically developed from the support verb **y* 'to say, do, be', when it receives the verbal ***-s** (i.e., ***-ys** → ***-is**, see §§4.2.5.7, 4.2.4.1). In other words, when **-is** is found after geminates/clusters, it historically goes back to causative ***-i-s** with epenthesis; otherwise, it is a reflex of the derived verb **is-* 'to make'. Still, the verb **is-* 'to make' may occur even in the environment of ***-i-s**. Leveling of the two is likely within individual languages, and non-transparent stem-final consonants may blur the distinction. Such a precaution is necessary because it is sometimes unclear whether some underlying stems comprise complex stem-final consonants (***CC**, ***cC**). For example, is PHEC [496] **raʔ-i-s-* 'to cook' derived from [431] **raʔ-* 'ripen, be ripe' with a single **ʔ** or does it include an underlying stem-final geminate **ʔʔ**?[105]

On the other hand, there are cases where only **-s** surfaces instead of the expected **-is** (or **-ees**) because the support verb has been deleted or fused as part of morpheme boundary alternation processes (§3.4.1.2). For instance, after a vowel-final stem, the support verb surfaces as **y** in Burji *battee* (adj) 'dirty, filthy' and it is retained in the denominalized *batte-ys-* 'to make dirty', but in *c'eejiy-* 'to bleed' the support verb becomes **i** in *c'eejis-* 'to bleed, make bleed' with a **-s** as a causative denominative following the historical pattern (or dropped and the current **i** is the stem final vowel, see §3.4.1). Such an unstable synchronic reality does not undermine the established historical scenario.[106]

Palatalized and Affricated ʃ/c

A palatalized causative **-ʃ** is found in three NHEC languages: Kambaata, Sidaama, and Gedeo. More than one scenario might have led to the innovation of the palatalized variant in these three languages. First, the palatalized **-ʃ** is driven by the unstable history of the post-alveolar ***ʃ** (§2.1.1.2)[107] as witnessed in individual language-internal and dialectal variation between the palatalized **ʃ** and alveolar **s** in verbal derivation – i.e., it is optional in Sidaama (**ʃ ~ s**) (Kawachi, 2007, pp. 52–53) and dialectal in Gedeo (highlanders prefer **ʃ** over **s**) (Eyob Kelemework, 2015, p. 182). Second, the palatalized **-ʃ** might have emerged as an allomorph in specific contexts: (a) following a stem-final **l** where its reflexes are mainly found in

[105] It remains to be determined how ***-is** (caus) and ***is** 'to make' were functionally distinguished (if at all) at the proto-level.
[106] Data source: Burji (Getahun Addo, 2017; Tesfaye Baye, 2015).
[107] For example, the stem final alveolar **-s** in Hadiyya and Kambaata *gos-* and the palatalized stem-final **-ʃ** in Gedeo and Sidaama *goʃ-* 'uproot'.

Kambaata (Treis, 2022, p. 12) and Gedeo (Eyob Kelemework, 2015, p. 94) (§3.1.1.3) and (b) as a result of an exceptional fusion process of the support verb *y* 'say, do' and **-s**. Instead of leaning towards a single scenario, a combination of such internal factors and contact has contributed to the diffusion of the palatalized **ʃ**, since it is also found in languages like Oromo.

Sometimes the palatalized **ʃʃ** of Kambaata, Sidaama, and Gedeo corresponds to geminated **-ss** in the remaining two HEC languages, i.e., Hadiyya and Burji, supporting the fusion interpretation, for example, Hadiyya [701] *giggiss-* 'to shake off' in comparison to Kambaata *giggiʃʃ-* 'shake off, let fall down' and Sidaama *giggiʃʃ-* 'tear down part of a house in order to renew it, snatch, kick off'.[108]

Similar to the nominal ***ʃ**, Gedeo affricates the palatalized verbal **ʃ** into **c** after stem-final **l**, as in *ful-* 'exit, get out' → *ful-c-* 'take out' (Eyob Kelemework, 2015, p. 189) in comparison to Sidaama *ful-* 'to exit' *fuʃʃ-* 'to move sth out, take off clothes' (Kawachi, 2007, pp. 55, 117).

Kambaata, Gedeo, and Sidaama apply the same distribution of the PHEC allomorphs to the recent palatalized **ʃ** variant. Thus, the same stem might receive either **-s** or **-ʃ**, depending on the language and sometimes in the same language (i.e., Sidaama or Gedeo).[109]

[108] The reflexes of the other two languages show completely different processes: metathesis of **-s** in Burji *jisk-*, *jig-* 'to demolish/destroy', and epenthesis or change of **y** to **i** in Gedeo *giggis-* 'stutter'.
[109] Data source: Hudson (1989).

	Verbal -s	**Gloss**	**Verbal -ʃ**	**Gloss**
Burji	**huus-**	'fill'	**-**	**-**
Had	**-**	-	**wanʃ-**[110]	'fill'
Kam	**-**	-	**wiinʃ-, yiinʃ-**	'fill'
Sid	**woʔm-is-**	'fill'	**wonʃ-**	'fill'
Ged	**-**	-	**wanʃ-**	'fill'
Had	**iibb-is-**	'to heat'	**-**	-
Kam	**iibb-iis-**	'to heat'	**-**	-
Sid	**-**	-	**iibb-iʃ-**	'to heat'
Ged	**-**	-	**eebb-iʃ-**	'to heat'
Had	**haraass-** ← **haraar-s-**	'widen'	**-**	-
Kam	**haraars-**	'widen'	**-**	-
Sid	**-**	-	**halaʔl-iʃ-**	'widen'
Sid	**-**	-	**dur-iʃ-**	'fry, roast'
Ged	**duʔris-**	'fry, roast'	**-**	-

Such varied patterns prove the recent suffixation of the palatalized **ʃ** as an independent verbal derivation suffix in Kambaata, Sidaama, and Gedeo, competing with the older **s**.

Alveolar s and Palatalized ʃ

In the double causative formation caus1-caus2, caus2 always follows the *y* of the support verb, regardless of the number of preceding causatives; i.e., caus2 requires a support verb.[111] Therefore, the common HEC double causative **-siis** is composed of ***-s** (caus1) followed by an epenthetic vowel **-i-** to separate it from the following ***ys** sequence of caus2, i.e., ***-s-i-y-s** → **-siis**. This hypothesis explains the double vowel of the double causative **-siis** via the fusion of the **y** and the following epenthetic **i**.[112]

[110] The presence of **ʃ** in Hadiyya *wan-ʃ-* (tr) 'fill' is an exception, as Hadiyya does not generally exhibit **ʃ**.
[111] The reason caus2 of the double causative requires the support verb *y* is that it is obligatorily attached to a verb root. This is supported by the need of the denominatives for a support verb (**y** or **ʔ**) to receive the suffix **-s** or any denominative (§4.2.5).
[112] Epenthesis as a sole hypothesis fails to account for the reason the single epenthetic vowel becomes double **-ii-** in the double causative **-siis** found in all HEC languages.

There is a variant of the double causative **-ʃiis** with palatalized **-ʃ-** in caus1 and alveolar **-s** in caus2 in the three languages where the palatalized causative is attested, i.e., Kambaata, Sidaama, and Gedeo. Therefore, **-ʃiis** is analyzed as a palatalized caus1 **ʃ** followed by an alveolar **s** in caus2 ***ys** → **is**, i.e., **-ʃ-i-y-s-** yielding **-ʃiis-**. The palatalized **-ʃ** is always the first member (caus1), probably because it was conditioned by the stem-final consonant, so it remains closer to the stem. Then a later **y** and a following causative **-s** were added, triggering **-i-** epenthesis before the **ys** sequence with a subsequent fusion of **-iy-** into double **-ii-**. The epenthesis in this case is motivated by the following **ys** sequence to avoid three consonant sequences in the double causative.[113]

Additionally, there is an optional palatalization of caus2 in double causative, but restricted to double causative where caus1 is a palatalized **ʃ**. Therefore, Sidaama and Gedeo allomorphs **-ʃiis** ~ **-ʃiiʃ** can be explained by arguing that **-ʃiiʃ** includes an optional obstruent harmony of the **-s** of caus2 to the palatalized **ʃ-** of caus1 in the allomorph **-ʃiis** (fn. 36). It must be the assimilation of caus2 **s** to caus1 **ʃ** because there is ***-ʃiis** (← ***-ʃ-i-y-s**) but no attestation of **ʃ** in caus2, i.e., there is no ***-siiʃ-**.

Contrary to the regular ***is**, the **ys** sequence can yield geminate **ss** in Burji and Hadiyya. Therefore, the double causative **-siss** is only attested in those two languages, suggesting a fused caus2 **-ss** composed of **y** and **s**. The variant **-siss** is a recent innovation in the two languages, because it is likely ***-s** followed by the support verb [388] **y* 'to say, do, be', and another causative ***-s**, which makes **-siss** different from the old double causative form ***-siis**, although the components are the same.

In summary, the verbal derivation ***-s** is historically realized as **-s** except after geminates and clusters, where it is preceded by epenthesis and becomes ***-is**. The derived verb form **is* 'to make' is one of the realizations of the support verb **y* 'to say, do, be' followed by the verbal **-s** after the change of **y** into **-i**, and is also used in verbal derivation. Kambaata, Sidaama, and Gedeo innovated a palatalized variant **ʃ**, which is affricated to **c** in Gedeo after **l**. Both the palatalized and alveolar geminates **ʃʃ**/**ss** involve a fusion process with a preceding consonant, including fusion with the **y** of a preceding support verb. The HEC double causative comprises either an alveolar **-s** or a palatalized **-ʃ**,

[113] There are instances of multiple causative suffixes, e.g., Gedeo, Kambaata, Burji **-sisiis** and Gedeo, Sidaama and Kambaata **-ʃisiis** where there is a single **-i-** before caus2, so there may be no epenthesis inserted.

followed by epenthetic **-i-** and a complex form **-is/-ys**, yielding either **-siis/-siss** or **-ʃiis**.

Stem-final s

Other than the verbal **s**, the stem-final **s** is also palatalized to **ʃ** in NHEC through a fusion process triggered by the addition of the derivation suffix ***-ʃ** (§§4.1.1.1, 4.1.2). For example:

> PNHEC [406] **wees-*, **wees-ʃ-* (sgv) 'ensete tree (Ensete ventricosum)' yields Kam *weesi-ta*, *weeʃʃu,* and Sid *weese*, *weeʃʃo,*
>
> PHEC [174] **hamas-* 'snake' → Burji *hamasi*, and PNHEC **hamas-*, **hamas-ʃ-* (sgv) 'snake' → Had *hamaʃʃicco*, *hamaʃʃa*, Kam *hamasu* 'roundworm (askarias)', Sid *hamaʃʃo*.

In [406], both the ***s** and palatalized **ʃʃ** are preserved. However, in the second example [174], only Burji and Kambaata retained the stem-final **s**, which has changed into **ʃʃ** in Hadiyya and Sidaama. The derivational morpheme ***-ʃ** is the source of the palatalization and gemination (***s+ʃ** → **ʃʃ**) in [406] **wees-ʃ-* (sgv) 'ensete tree (Ensete ventricosum)', and [174] **hamas-ʃ-* (sgv) 'snake'.

3.1.2.3 Stem-final n

The alveo-palatal nasal **ɲ** is rare in the HEC languages (Bender, 2020, pp. 116–117). It typically occurs in loanwords (e.g., Burji; Sasse, 1982, p. 16), in onomatopoeic forms (e.g., Sidaama; Indrias Xamiso et al., 2007, p. 632; Anbessa Teferra, 2000, p. 13), or as an allophone before palatals (e.g., Burji; Hudson, 1976, p. 248). Tesfaye Baye (2015, p. 39) provides an example of the latter in Burji:

> *k'un-* 'pound' → *k'un-t-ng* → *k'uɲcingi̯* 'they pounded (and)'

However, Tesfaye Baye (2015) provides examples of non-palatalized **n** before a palatal **c** in the third-person feminine agreement for the same word *kúnci̯*. Another non-palatalized **n** is found within the deverbative suffix *-aanco* in *k'un-aanco* 'grinding stone'. If these are free variants, they constitute a counterexample to Hudson's rule. Of course, other (non-palatal) suffixes also have no impact on the stem-final **-n** of *k'un-* 'pound', e.g., the third-person feminine jussive suffix *-tooni* (3sg.f.jussive) → *k'undoóni̯* 'let her pound' and nominal derivation suffixes *k'un-toó* 'pounder'.

In summary, alveolar consonants undergo productive changes at stem/suffix boundaries throughout HEC. Full assimilation of the pronominal suffix **-t** to a preceding consonant is an NHEC innovation, systematically attested in Hadiyya, Kambaata, Sidaama, and Gedeo, but not in Burji. The affrication of the nominal derivational suffix ***-ʃ** is found in all HEC languages, and the reflexes of post-alveolar ***-ʃ** in both Burji and NHEC support its reconstruction to PHEC. Palatalization of **t** → **c**/**ʃ**, and more marginally **n** → **ɲ**, is attested in multiple contexts. These palatalization patterns provide evidence for reconstructing the causative suffix ***-s**; by contrast, the palatalized Kambaata, Sidaama, and Gedeo variant **-ʃ** is likely a later innovation for which the conditioning still requires further investigation. Taken together, these boundary processes refine the cognate-identification framework established in §1.5, as superficially irregular forms become transparent once their derivational history is recognized.

3.2 Velars

Three morpheme boundary alternation processes affect the velars **k** (and **g**): palatalization, lenition, and glottalization. This section treats palatalization and lenition. Glottalization, which results from fusion with a following glottal stop, is discussed in §3.3.1.

3.2.1 Palatalization

In addition to the palatalization of the alveolar **t** to **c**, **ʃ** (§3.1.2.1), Burji palatalizes the velar ***k** to **ʃ** before the high front vowel **i** (or older Nominative ***y** §4.1.5.2), including in pronominal morphology.

k + i → ʃi

	PHEC	k+i	Nominative	t+a	Accusative	Gloss
[253]	**ki-ʔ-*	*ʃíccị*	(2sg.poss.f)	*ʃíttạ*	(2sg.poss.f)	‘your’ (2sg.poss)
[187]	**kin-ʔ-*	*ʃíncị*	(2pl.poss.f)	*ʃíntạ*	(2pl.poss.f)	‘your’ (2pl.poss)

A marginal case of possible palatalization **g** → **j** is found in two examples in Burji. The following comparative sets list examples from Burji, Hadiyya, Kambaata, and Oromo:

[18]	g	j	Gloss
Burji	-	**c’eeji**	‘blood’
Had	**t’iiga**	-	‘blood’
Kam	**k’egu/k’egi**	-	‘blood’
Oro	**d’iiga**	-	‘blood’

It is possible that the palatalized **j** triggered the obstruent harmony, hence the initial ***t'*** became **c'** in Burji. The same harmony process is responsible for changing **t'** to **k'** in Kambaata, but via assimilation to the velar **g** (fn. 36).

[708]	**g**	**j**	**Gloss**
Burji	**moorga**	**moorjoo**	'thief'
Kam	**moga?-aancu**	**-**	'thief'

The Burji variant with **j** might have been followed by **i** before the current terminal **oo**. The two apparently differ in number and gender: *moorga* 'thieves' (f) and *moorjoo* 'thief'(m) (Getahun Addo, 2017, p. 677).

The distribution of the palatalized alveolars and velars in Burji can be summarized as:

		initial	non-initial	
***t**	→	**c**	**-**	**#_i, C_i**
***t**	→	**-**	**ʃ**	**V_i**
***k**	→	**ʃ**	**-**	**_i**
***g**	→	**-**	**j**	**_i**

Palatalization underscores Burji's innovative phonological trajectory in a small set of lexemes. The absence of comparable developments in the other HEC languages highlights Burji's divergence and raises questions about possible areal influences (Oromo and EOT palatalization).

3.2.2 Lenition

Intervocalic or non-initial single ***k** → **h*** occurs in Burji (Sasse, 1982, p. 18), Sidaama (Kawachi, 2007, p. 47; Anbessa Teferra, 2000, p. 26; Hudson, 1989, p. 7), and Kambaata (Treis, 2008, p. 32, 63). This lenition is also relevant for alternations involving **h**, including the Kambaata denominative **-eeh** discussed in §4.2.5.4. Historically, the instances of corresponding **k**, **h**, **ø** reflexes must be attributed to a lenition process similar to Sidaama *duk-* 'carry' → *duhí* (perf) 'he carried' in comparison to Gedeo *duk-* 'carry'.

Vk + V → h

If there are any uniform cases of an older ***k** → **h***, they must have been mostly within shared HEC stems and, thus, eliminated via the subsequent loss of non-initial post-vocalic/intervocalic ***h***, e.g., PEC [299] **zakn-, *zakan-*

‘elephant’ via **zahan-* → PNHEC **zaan-* ‘elephant’. Here are examples of corresponding **k**, **h**, **ø** reflexes:

[125]	**k**	**Lenited h**	**Gloss**
PHEC	***buk-**	**-**	‘dust’
Burji	**-**	**buhoo**	‘dust’
PNHEC	***buk-**	**-**	‘dust’
Kam	**-**	**buhu-ta**	‘dust’
Sid	**buko**	**-**	‘dust’
Ged	**buko**	**-**	‘dust’

Hadiyya has *bucca* ‘dust, soil, ground’, with a possible assimilation of stem-final **h** and singulative **c**.

[673]	**k**	**Lenited h**	**Gloss**
PHEC	***ik-**	**-**	‘to be(come), happen’
Burji	**-**	**ih-**	‘to be; become’
PNHEC	***ik-, *ikk-**	**-**	‘to be(come), happen’
Had	**ikkoʔo**	**ih-**	‘to be, become, be enough’
Kam	**ikk**	**ih-**	‘be, suffice’
Sid	**ʔikk-**	**-**	‘become, be, happen’

It is unclear if Gedeo *kada* ‘become, be, happen’ (with a puzzling **d**) is related to the rest of set [673].

[466]	**k**	**Lenited h**	**Lost ø**	**Gloss**
PHEC	***diwk-**	***diwh-**		‘empty’
Burji	**-**	**-**	**duww-**	‘empty and deserted house’
PNHEC	***diwk-**	***diwh-**	**-**	‘empty’
Had	**-**	**diiha**	**-**	‘empty’
Kam	**-**	**diiha**	**-**	‘empty’
Sid				
Ged	**duwwika**	**-**	**duww-**	‘empty, bare, naked’

See Omoro *duwwaa* ‘empty’.

[118]	**k**	**Lenited h**	**Lost ø**	**Gloss**
PHEC	***zaak-**	**-**	**-**	‘to swim, to spil’
Burji	**daak-**	**daaha**	**-**	‘swim’
PNHEC	***zaaz-, *zaak-**	**-**	**-**	‘to swim, to spil’
Had	**-**	**-**	**(daad-)**	‘spill, flow’
Kam	**-**	**-**	**(zaaz-)**	‘spill, flow’
Sid	**daak-i**	**daah-i**	**-**	‘swim’
Ged	**daak-**	**-**	**-**	‘swim’

The glosses of Kambaata and Hadiyya parallels make them doubtful cognates. See Oromo *daaka* ‘swim’.

Overall, there is a tendency for Gedeo and Sidaama to retain the ***k** while lenition is more prevalent in Burji, Hadiyya, and Kambaata.

Additional examples include: PHEC [445] **bork-* ‘to be impure’, PHEC [723] **sek-* → **seh-* ‘adorn, decorate’, [15] **bezz-k-* ‘star(s)’, [258] **maak-*, **ma~mmaak-* → PNHEC **maak-* ‘say in a proverb, tell a story’.

Same words are sometimes transcribed with **h**, **k**, and **ø** depending on the source, e.g., Gedeo *duwwi* (Wedekind, 2008, p. 45), *duwwika* (Hudson, 1989, p. 56), and Sidaama [15] *beeddahe* (Yri & Pepper, 2019), *beddakko* (Hudson, 1989, p. 142) ‘star(s)’. The geminated **kk** might have been triggered by vowel-initial morphemes but also morphemes with an underlying initial consonant that only manifests in the geminated stem-final **k** via the regular morpheme boundary alternation rule of assimilation, e.g., Sidaama *beddahe* (pl.f) and *beeddakko* (f) ‘star’ (§3.1.1). Older gemination is attested in PHEC [51] **lokk-*, **lekk-* ‘foot, leg’ (see Sasse, 1982, p. 136), which otherwise has a single final **k** in the rest of EC. However, it is difficult to specify the quality of such a fully assimilated underlying consonant.

In a cluster/gemination, the distinction between **k** and **h** is neutralized in favor of **k**, at least in Sidaama (Yri & Pepper, 2019) and Kambaata (Treis, 2008, p. 32). In Hadiyya, Tadesse Sibamo (2015, p. 48) claims “…an alternation of the voiceless velar consonant /k/ with the glottal stop /ʔ/. Such alternation takes place when the glottal stop, /ʔ/, occurs as a final consonant of the noun stem”, e.g.:

biimbéʔe (citation form), *biimbekk-ícc* (sg) ‘a/the fly’[114]

However, there is comparative evidence for ***k** → **h** but not ***k** → **ʔ**. Thus, Hadiyya **k** might be part of the stem but eliminated via the lenition of ***k** intervocalically and the loss of single non-initial ***h**, then replaced by an epenthetic **-ʔ-** in the citation form: ***VkᵢV** → ***VhᵢV** → **VøᵢV** → **VʔV** (§3.3.4). Based on the historical evidence, the derived form has an underlying ***k** which surfaces when geminated ***VkV** → ***VkkV** in Hadiyya, Kambaata, and Sidaama.

According to the HEC grammars, the stem-final fricative **h** is dropped (see Wedekind, 1980, p. 137). In Sidaama, **h** is elided when preceded by a vowel (Anbessa Teferra, 2000, p. 27). Similarly, Gedeo deletes stem-final **h**, which sometimes results in a compensatory doubling of the preceding vowel, e.g., Gedeo *doh-* ‘rotate’ → *doo-t-e* (3sg.f-sg.perf), *doo-n-e* (3pl-sg.perf) (Eyob Kelemework, 2015, pp. 83–85). By contrast, in Burji and Hadiyya, **h** is banned in syllable codas (Tesfaye Baye, 2015, p. 61; Tadesse Sibamo, 2015, p. 34).

However, historically, the reflexes of PHEC ***h** have been lost post-vocalically. Thus, the remaining attestations of post-vocalic **h** are either later lenited ***k** → **h** or reflexes of PHEC ***ħ** (§2.1.5.3).

The processes of **k** and **g** variation (e.g., in Burji (Tesfaye Baye, 2015, p. 79; Sasse, 1982, p. 18)), lenition of the voiceless velar **k**, loss of lenited **h**, and insertion of **-y-/-w-** (§3.4.1.2) should be analyzed as interrelated processes that feed into one another. For instance, Burji’s alternation of **g** and **w** (Tesfaye Baye, 2015, pp. 78–79) before **u** (Sasse, 1982, p. 18) in some words could be interpreted as going back to an older ***k**, which is lenited and dropped in non-initial contexts, after which **-w-** is inserted. Burji alternations are due to dialectal variations.

The velar consonants **k** and **g** exhibit two distinct alternation patterns. Palatalization of **k** → **ʃ** before front vowels is restricted to Burji, reflecting a Burji-specific innovation. Intervocalic lenition (**k** → **h** (→ **ø**)) is attested across HEC, though with variable degrees of lenition in different daughter languages. The interaction of these patterns with nominal and verbal morphology (Ch. 4) provides additional diagnostic evidence for identifying cognate sets where surface forms have diverged significantly.

[114] cf. *ʔasseʔ-* ‘send’ → *ʔassé-cc* ‘message’ (Tadesse Sibamo, 2015, p. 48).

The next section turns to the glottal stop itself, showing how suffixal **ʔ** drives regular fusion outcomes and how it interacts with hiatus repair.

3.3 Glottal Stop

The glottal stop **ʔ** occupies a central role in HEC verbal and nominal morphology. As a derivational suffix, it marks reflexive/middle voice, functions in denominative constructions, and serves as a support verb *ʔ* 'to do'. All five HEC languages preserve the reflexive suffix **ʔ** (Burji (Tesfaye Baye, 2015, p. 183), Hadiyya (Tadesse Sibamo, 2015, p. 77), Kambaata (Treis, 2008, p. 217), Sidaama (Kawachi, 2007, p. 335), and Gedeo (Eyob Kelemework, 2015, p. 190)). This regularity across the family provides secure evidence for the reconstruction of PHEC **-ʔ** as a derivational suffix. The morphological roles of **ʔ** are treated in Chapter 4: reflexive (§4.2.3.1), verbalizer (§4.2.5.1), and support verb (§4.2.4.2). The following examples illustrate how **ʔ** is utilized for diverse and often overlapping verbal derivations:

> PHEC [36] **gam-* 'to bite' → Burji *gam-*; PNHEC **gam-ʔ-* → **gaʔm-* → Ged, Sid, Had, Kam *gaʔm-*;
>
> PHEC [256] **ham ʔ* 'to slander' → Burji *ham-ad'-*; PNHEC **ham ʔ* 'to slander' → Kam and Sid *heʔm*, Had *ham* (cf. Ged *ham-at* ← Oromo *hammaat-* (adj) 'bad off, unwell');
>
> PHEC [340] **gag~gal-* → **ga-n-gal-* 'to roll' → Burji *gaangiy-*, *gangald'-*; PNHEC **gog~gol-* → **go-n-gol-* 'to roll' → Had *gunguuʔl-*, Ged *goggoʔma*, Sid *gongoʔm-*, *gongomi*.

A fossilized nominal suffix **-ʔ** is proposed with strong evidence in Hadiyya and Gedeo, which is found replaced by other nominal suffixes and as potential remnants in the rest of HEC (§4.1.3). The second nominal **ʔ** is found as a genitive marker in Hadiyya (§4.1.4.1).

The present section addresses the morpheme boundary alternations of **ʔ** when it concatenates with stem-final consonants or another suffix. In NHEC, sequences involving the glottal stop trigger different resolution strategies depending on the preceding segment: fusion with obstruents (§3.3.1), metathesis with sonorants (§3.3.2), deletion in stem-final position (§3.3.3), and epenthetic insertion intervocalically (§3.3.4). Because fusion and metathesis are confined to NHEC, they constitute important evidence for NHEC as a subgroup (see Ch. 5).

3.3.1 Fusion

When the derivational suffix **ʔ** attaches to a stem ending in an obstruent, the two segments fuse, producing a glottalized (ejective) geminated consonant (**C'C'**). Stem-final sonorants behave differently in general (see §3.3.2 on sonorant + ʔ metathesis); however, an important Sidaama/Gedeo pattern suggests a distinct contact-conditioned outcome for stem-final **-r** (introduced in §3.3.1.5).

$C_{[-son]}$ + ʔ → C'C'

The resulting **C'C'** is phonetically distinct from inherited non-pulmonics (§1.5.2): fusion-derived glottalized consonants are always geminated post-vocalically and arise transparently from morpheme concatenation, whereas non-pulmonics are inherited and may occur as singletons post-vocalically and lack synchronic derivational conditioning.

Crucially, Burji does not exhibit such a fusion. The process is therefore reconstructed as a PNHEC innovation, providing morphophonological evidence for the NHEC subgroup. Burji instead shows the borrowed **-d'** reflexive variant assimilating to the stem-final consonant through a different pattern (§4.2.3.4). The subsections below detail the fusion outcomes by stem-final consonant: labial ***b** (§3.3.1.1), alveolars ***t, *d** (§3.3.1.2), velars ***k, *g** (§3.3.1.3), and post-alveolar ***ʃ** and alveolar ***s** (§3.3.1.4).

All NHEC languages (Gedeo, Sidaama, Hadiyya, Kambaata) show the pattern **$C_{[-son]}$ + ʔ → C'C'**, with the input obstruent devoicing and the output emerging as a glottalized geminated consonant.[115]

[115] Data sources: Gedeo (Eyob Kelemework, 2015, pp. 190–191, 72–73), Sidaama (Kawachi, 2007, pp. 57–58, 35), Hadiyya (Tadesse Sibamo, 2015, p. 133), and Kambaata (Treis, 2008, pp. 243, 269, 242, 267; Treis, p.c.).

	Glottalization and gemination	**Example**
Ged	**b, d, g, j + ʔ → p'p', t't', k'k', c'c'**	*gub-* 'burn' → *gup'p'-e-* *abid-* 'catch' → *abit't'-*
Sid	**b, d/t, k/g, j + ʔ → p'p', t't', k'k', c'c'**	*gob-* 'to sew' → *gop'p'-* *daak-* 'to grind' → *daak'k'-*
Had	**b(b), d/t, ʃ/j + ʔ → p'p', t't', c'c'**	*gub-* 'to drink' → *gup'p'-* *ʔamad-* 'catch' → *ʔamat't'-* *ʔijaaj-* 'to order' → *ʔijaac'c'-*
Kam	**b, d/t, k/g, ʃ/z/s + ʔ → p'p', t't', k'k', c'c'**	*k'orab* 'wait' → *k'orap'p'-* *agud-* 'resemble' → *agut't'-* *dag-* 'find' → *dak'k'-* *kuʃu* 'choke' → *kuc'c'u*
Burji	**b**, **f + d' → d'd'**	*rab-* 'paint' → *rad'd'-* *ʃaf-* 'carve' → *ʃad'd'-*

Eyob Kelemework (2015, pp. 89–90) proposes that the **-d'** variant of the reflexive/middle suffix triggers fusion in Gedeo. However, phonetic and comparative evidence contradict this analysis. If the voiced implosive **d'** were the trigger, we would expect the output to retain voicing. Yet fusion consistently produces voiceless glottalized geminates (**C'C'$_{[-voice]}$**).[116] The devoicing effect is better explained by the voiceless glottal stop **ʔ** as the fusing segment, consistent with the pattern across NHEC. Gedeo, therefore, patterns with Sidaama, Hadiyya, and Kambaata in using **-ʔ** for fusion, not **d'**. The current outcome suggests that the voiceless glottal stop **ʔ** is the variant used in the fusion process, which helps to account for the devoicing effect along with other reasons listed in §4.2.3.1. Kawachi's (2007) analysis of Sidaama does not specify which variant of the suffix (**ʔ** or **d'**) drives the glottalization and gemination process. In Kambaata, while Treis (2008) does not explicitly describe all instances of fused **C$_i$+ʔ**, except **d + ʔ → t't'**, the pattern can be inferred from available examples.

Although not listed in the grammar, Hadiyya attested the process **k + ʔ → k'(k')** at least in fossilized forms. However, its scarcity reflects the historical loss of non-initial **-k** through lenition and subsequent deletion (§3.2.2). Hadiyya and Kambaata share a distinctive reflexive suffix **-k'k'** not found elsewhere in HEC. This form likely derives from ***-k-ʔ** via fusion (**k + ʔ → k'k'**). The parallel with Omotic **-k'** (middle voice marker), attested in Bench (Rapold, p.c.), raises the question of whether this reflects an old areal feature or

[116] That is, regardless of the fact that the phonetic realization of **d'** is debated in the East Cushitic literature (Hudson, 1976, p. 248; Bender, 2020, pp. 116–117; Treis, 2008, p. 27; Yri & Pepper, 2019). More generally, voicing is common but not defining for implosives.

independent development. In Hadiyya, the scarcity of **k** + **ʔ** → **k'(k')** examples reflects the historical lenition and loss of non-initial **-k** (§3.2.2).

Burji does not show fusion of stem-final obstruents with a following **ʔ**, confirming that this process is a PNHEC innovation. Instead, Burji employs the borrowed **-d'** variant of the reflexive suffix, which assimilates to the stem-final consonant in a distinct pattern, e.g., *rab-* 'paint' → *rad'd'-* 'paint oneself'.(§4.2.3.4):

Burji: **C** + **d'** → **d'd'**

Exceptions exist where **-d'** attaches without assimilation, e.g., *gaaf-* 'grind' → *gaafd'* 'grind for oneself' (Tesfaye Baye, 2015, pp. 182–183). Glottalized consonants that do appear in Burji are either inherited non-pulmonics, loanwords, or morphologically exceptional forms. The absence of productive $\mathbf{C_{[-son]}}$ + **ʔ** → **C'C'** fusion in Burji supports the reconstruction of fusion as a shared innovation of PNHEC.

The following sections discuss how the glottal stop **ʔ** has contributed to the emergence of the glottalized consonants across the NHEC languages through fusion processes.

3.3.1.1 Stem-final b

Stem-final ***b** undergoes two distinct processes depending on the following suffix: (a) glottalization (fusion) triggered by derivational ***ʔ** (reflexive, denominative), (b) fricativization triggered by suffixes containing ***t** or ***s** (e.g., certain denominative and causative suffixes).[117]

***b** + ***ʔ** → glottalized **p'(p')**

***b** + ***-t**, ***-s** → ***ff** → **f**

The intermediate stages of fricativization may include geminate ***ff**, subsequently simplified to **f**.

[117] The causative ***-s** as a condition could be challenged by a synchronic morphophonemic process in which **b+s** → Burji, Kambaata, Gedeo **ss**. The denominative ***-t** should also be preceded by a vowel (see §4.2.5.3) to prevent it from fusing with the stem-final ***b**. However, the vowels of the denominatives originate from the support verb, which can be lost between consonants, allowing the fusion.

For example, Kambaata has *huf-* ‘comb’ → *hup’p’-* ‘comb oneself’ and *t’oof-* ‘finish’, *t’oop’p’-* (Treis, 2022; Treis, p.c.). The following lists include more examples of the two processes:

[29]	**b**	**Glottalized p’(p’)**	**Fricativized f**	**Gloss**
PHEC	***k’ab-**	**-**	**-**	‘to hold, seize’
Burji	**-**	**-**	**k’af-**	‘to have, possess’
PNHEC	***ab-**	**-**	**-**	‘to hold, seize’
Kam	**-**	**ap’p’-**	**-**	‘catch fire, hold’
Ged	**abid-**	**-**	**-**	‘hold, seize’

Many of the examples with such glottalized consonants count as fossilized evidence – they do not have a stem with a non-glottalized reflex, but the comparative evidence provides such evidence. Cognate set [29] is one of these examples, with a corresponding retained **b** in Gedeo and the EC cognates.

The Burji **f** reflex of ***b** in *k’af-* ‘have’ is described by Sasse (1982, p. 123) as irregular, although he notes cognate Dirayta *k’af-naad’-* as an extended stem of *k’ap-* ‘have’.[118] The **f** reflex is not limited to Burji and is found in more than one example in all HEC languages. Its restricted occurrence on verbs proves the proposed morphophonological trigger of affrication.

[186]	**b**	**Glottalized p’(p’)**	**Fricativized f**	**Gloss**
PHEC	***c’ub-**	**-**	**-**	‘dip in, immerse’
Burji	**c’uub-**	**-**	**-**	‘immerse, to dip’
PNHEC	***c’ub-**	**-**	**-**	‘to rinse, immerse’
Had	**-**	**-**	**c’ufc’ufaaʔ-**	‘to wash’
Kam	**-**	**-**	**c’uf aʔ**	‘rinse, immerse’
Ged	**-**	**c’uup’-**	**-**	‘dip (in), immerse’

The Sidaama cognate *c’uuʔ-* ‘baptise, immerse’ has dropped the labial completely, which is expected in intervocalic position, hence Indrias Xa’miso et al. (2007, p. 97) transcribe it as *cuua* ‘to immerse, dip in water, baptize’. The additional **aaʔ-** in Hadiyya and **aʔ** in Kambaata are the reflexes of the support verb **ʔ* ‘to do’ used to derive the denominal. Based on the above synchronic pattern, single **C’** as in Gedeo *c’uup’-* ‘dip (in), immerse’ can be labeled as contact-influence on the basis of Oromo *c’uup’-* ‘dip, baptize’

[118] Dirayta **p** is realized as **f** before any consonant: e.g., *k’ap-* ‘catch’ followed by **-te** is realized as *káf-té* ‘she caught’ (Black, 1974, p. 70).

because the expected result of the fused glottalized consonant is geminated **C'C'**. Otherwise, obstruent (glottalization) harmony of **b** with the initial **c'** (fn. 36) is an alternative plausible explanation.

[430]	**b**	**Glottalized p'(p')**	**Fricativized f**	**Gloss**
PHEC	***gilub ʔ, *gulib ʔ**	**-**	**-**	'to kneel'
Burji	**gilb-is-ad'-**	**-**	**-**	'kneel'
PNHEC	***gulub ʔ**	**-**	**-**	'to kneel'
Had	**-**	**gurup'p'-**	**-**	'kneel'
Kam	**-**	**gulup'p'-**	**-**	'kneel'
Sid	**-**	**gulup'p'-**	**-**	'kneel'
Ged	**-**	**-**	**guluff-at-**	'kneel'

This verb is derived from the noun [4] **gulib-, *gilub-* 'knee' via the denominative suffixes. The one responsible for glottalization in Hadiyya, Kambaata, and Sidaama is **-ʔ** and the one responsible for the affrication in Gedeo is **-t**. Burji has more than one derivation suffix, namely **-is** and **-ad'**, but the latter is borrowed from Oromo.

Overall, glottalization is more common than fricativization. The cases of fricativized ***b** → **f** are confined to the verbs as seen in the above examples, proving that a verbal suffix has conditioned the change. The fusion of ***bt/*bs** is expected to produce a geminate **ff**, but no HEC language has retained ***ff**, so it is possible that an intermediate geminate ***ff** reduced to **f**.

On the nouns, the glottal stop that might have triggered the glottalization is problematic (§4.1.3), and there is no fricativization. Unlike the verbs, the glottalized consonants of the nouns are single.

[544]	**b**	**Glottalized p'(p')**	**Gloss**
PNHEC	***c'ub-**	**-**	'dagger, knife'
Had	**c'ubeʔe**	**-**	'dagger'
Kam	**-**	**c'up'i-ta**	'dagger'
Sid	**c'uube**	**-**	'dagger'
Ged	**-**	**c'uup'e**	'knife'

Like [186], the single **p'** in the Kambaata and Gedeo reflexes in [544] can be a result of obstruent (glottalization) harmony of **b** with the initial **c'** (fn. 36). The parallels Amharic *c'ube* and Oromo *c'uubee* 'dagger' do not show a glottalized **p'**, which undermines the possibility of explaining this through contact, but does not preclude the possibility that the non-glottalized stems were borrowed

from these sources. However, the harmony analysis does not apply to the next two examples.

[180]	**b**	**Glottalized p'(p')**	**Gloss**
PHEC	***utub-**	**-**	'pole, center-pole of house'
Burji	**utubaa**	**-**	'pole, center-pole of house'
PNHEC	***utub-**	**-**	'pole, center-pole of house'
Had	**utuba**	**-**	'pole, center-pole of house'
Kam	**utubu**	**-**	'pole, center-pole of house'
Sid	**ʔutuba**	**-**	'private property'
Ged	**utuba**	**utup'a**	'pole, center-pole of house', 'support, e.g. forked stick

[81]	**b**	**Glottalized p'(p')**	**Gloss**
PNHEC	***kob-ʔ-**	**-**	'shoes'
Had	**kobeʔe**	**-**	'shoes'
Sid	**-**	**kop'é**	'sandal, shoe'
Ged	**-**	**kop'e**	'old type of shoes'

Oromo has both *kop'-ee* and *kob-ee* 'sandal'.[119]

Both the glottalization and fricativization result from the fusion process. Conforming with the current proposal, Sasse (1979, p. 14) assumes that most **p'** or **b'** reflexes can be explained by a hypothesis of secondary glottalization, unless they are due to borrowing. Sasse (1982, p. 16) also notes that **p** and **p'** occur only word-medially in Burji.[120] Hudson (1989, p. 12) describes Burji **b'** and **p'** as possibly idiolectal or dialectal variants, while Sasse (1979, p. 14) says they occur as irregular reflexes of ***b** in some languages.

According to Bender (2020, p. 121), the "…"f for p" phenomenon is strong in Cushitic". Could the change ***b** → HEC **f, (p')p'** alternatively be analyzed via an intermediate ***p**? The restricted non-initial environment where this change is attested in HEC supports the proposed morphophonological interpretation.

[119] The other common word for 'shoes' in NHEC is [81] *c'amma* 'shoes' from Amharic, found in all four languages.
[120] There is no attestation of **p**, while **p'** is only found in three loanwords in the Burji data used in this study.

3.3.1.2 Stem-final and Derivational t, d

The alveolar stops **t** and **d** are attested stem-finally and as derivational suffixes. When followed by the glottal stop, they undergo fusion in NHEC:

t, **d** + **ʔ** → **t't'**

The outcome is glottalized geminated **t't'**.

The following are the comparative data for the glottalization of stem-final and derivational **t**.

[605]	**t**	**Glottalized t'(t')**	**Gloss**
PHEC	***galat-**	**-**	'thanks, praise'
Burji	**galataa**	**-**	'thanks'
PNHEC	***galat-**	**-**	'thanks, praise'
Had	**galata**	**galat't'-**	'thanks', 'to thank, to praise'
Kam	**galata**	**galat't'-**	'thanks', 'to thank, to praise'
Sid	**galata**	**-**	'thanks'
Ged	**galata**	**-**	'thanks'

In cognate set [605], stem-final **t** undergoes fusion with **ʔ**, yielding **t't'** in Hadiyya and Kambaata in the derived verb. The non-glottalized variants and the rest of the HEC parallels are nouns. The noun stem is from Oromo *galata* 'thanks'.

[623]	**t**	**Glottalized t'(t')**	**Gloss**
PHEC	***oot-**	**-**	'fence'
Burji	**oota**	**-**	'enclosure, fence'
PNHEC	***oot-**	**-**	'fence'
Had	**oota**	**oot't'oro**	'fence', 'enclosure'
Kam	**oota-ta**	**-**	'fence'
Sid	**-**	**hut't'a**	'fence'

According to Hudson (1989, p. 57), Hadiyya [623] *oot't'oro* 'enclosure' is a compound of *oota-waro* 'fence-interior'. The glottalization might have occurred during the verbal derivation process, as in the Sidaama cognate (which is a noun and a verb).

[120]	**t, ø**	**Glottalized t'(t')**	**Gloss**
PHEC	***fil-ʔ-**	**-**	'to comb'
Burji	**fil-, sil-**	**-**	'to comb'
PNHEC	***fil-ʔ-**	**-**	'to comb'
Had	**-**	**fit'-**	'to comb, split, scatter'
Kam	**fiil-**	**-**	'unravel'
Sid	**-**	**fit't'-**	'comb, to edit a text'
Ged	**fil-**	**-**	'comb'

Both the **t** and glottal stops are possible concatenated suffixes. The single **t'** in Hadiyya might be a mistranscription.

[47]	**t, ø**	**Glottalized t'(t')**	**Gloss**
PHEC	***k'uf-**	**-**	'to cough'
Burji	**k'ufey-**	**-**	'to cough'
PNHEC	***kuf-**	**-**	'to cough'
Had	**-**	**kut't'**	'cough'
Ged	**k'ufa-aʔ-**	**-**	'cough'

The Hadiyya form [47] *kut't'-* 'cough' reflects the expected fusion **t** + **ʔ** → **t't'** (§4.2.3.2). Oromo *k'ufaʔa-* 'cough' is formally identical to Gedeo; however, the glottal stop here likely reflects a verbalizer ***ʔ**, as Burji shows the corresponding verbalizer ***y** in the cognate. Kambaata diverges, showing *kuc'c'-* 'cough' with **c'c'** rather than **t't'**. This suggests a different historical pathway, i.e., an underlying sibilant that fused (**s/ʃ** + **ʔ** → **c'c'**, see §3.3.1.4).

One example of stem-final **d** glottalization is the Gedeo stem [29] *abid-* 'hold/seize', mentioned above, which has a derived reflexive *abit't'-* 'keep/hold to oneself'.

Additional examples show singleton **t'** rather than geminated **t't'**. This likely reflects phonotactic constraints against sequences of three consonants, apocope, loans, or exceptions: PHEC [23] **hiir-* 'to untie' → Had *hirit'-* 'to shock', [362] **moor ʔ* 'to steal' → Had *maraat'-*, [87] **k'ot-* 'to dig' → Kam *k'ut'-* 'cut off, bite off', [94] **goraʔ-* 'to slaughter (cattle)' → Kam *gorat'-* 'slaughter (cattle)'.

The morphological evidence for glottalized **t't'** as a derivational suffix in Hadiyya and Kambaata is further documented in Chapter 4 under the reflexive (§4.2.3.3) and reflexive denominative (§4.2.5.5).

3.3.1.3 Stem-final k, g

Stem-final velar stops **k** and **g** also participate in fusion when followed by the glottal stop. Evidence for this process is limited because non-initial ***k** is subject to lenition (§3.2.2) and elision (§3.2.1) in non-initial environments. The attested examples involve consonant clusters (**Ck**), geminate velars (**kk**), or remnants of **-k** suffixes (§§4.2.3.2, 4.2.5.4).

***k, g + ʔ → k'(k')**

As with other fusion processes, this is a PNHEC innovation; Burji does not show the fused outcome.

[445]	k/g	Glottalized k'(k')	Gloss
PHEC	***bork-**	**-**	'to be impure'
Burji	**borkaawa**	**-**	'to become dirty'
PNHEC	***bork-**	**-**	'to be impure'
Had	**-**	**bork'-**	'be impure'
Kam	**-**	**bork'-**	'be impure'
Ged	**booroʔ-**	**-**	'be impure'

In Set [445], the verb is likely derived from a nominal stem via the verbalizer **ʔ**, as supported by the evidence of the noun stem in Burji *borkee* 'dirt, filth, stain', with an unfused velar.

[267]	k/g	Glottalized k'(k')	Gloss
PHEC	***t'ulak-**	**-**	'claw, fingernail'
Burji	**t'unga**	**-**	'finger'
PNHEC	***t'ulak-**	**-**	'claw, fingernail'
Had	**-**	**t'urank'a**	'claw, fingernail'
Kam	**t'ulunga-ta**	**t'ulunk'a-ta**	'claw, fingernail'
Sid	**-**	**c'ulunk'a**	'claw, fingernail'

The existence of both glottalized **k'** and non-glottalized **g** variants in Kambaata fits the obstruent (glottalization) harmony explanation (fn. 36) in this example in all three languages. In addition to *t'urank'a*, Hadiyya has *t'uraʔa* without the **k** or **k'**. This variant and Gedeo [445] *booroʔ-* 'be impure' above might be due to the lenition of velar **k** intervocalically (§3.2.2).

[321]	**k/g**	**Glottalized k'(k')**	**Gloss**
PHEC	***bukk-**	**-**	'to pluck'
Burji	**bukka**	**-**	'to cut into small pieces'
PNHEC	***bukk-, *bukk-ʔ-**	**-**	'to pluck'
Kam	**-**	**buk'k'-is-**	'pick, uproot'
Sid	**-**	**buk'k'-is-**	'weed, pluck'
Ged	**bukkassa**	**-**	'gather, collect'

See the identical parallel in Oromo *buk'k'-is-* 'weed, pluck' in comparison to Kambaata and Hadiyya.

[222]	**k/g**	**Glottalized k'(k')**	**Gloss**
PHEC	***t'ok-**	**-**	'to burst'
Burji	**t'okk-**	**-**	'to leak'
PNHEC	***t'ook-**	**-**	'to burst'
Had	**-**	**t'ook'-**	'burst'
Kam	**-**	**t'ook'-**	'burst'
Sid	**t'ook-, d'ook-**	**-**	'burst, crack'

Again, the glottalization in Kambaata and Hadiyya could be due to the obstruent (glottalization) harmony explanation (fn. 36), which is more plausible in cases like this where both consonants are voiceless.

[39]	**k/g**	**Glottalized k'(k')**	**Gloss**
PHEC	***suk-ʔ-**	**-**	'to spin'
Burji	**sukk-**	**-**	'to tie tightly, to knead'
PNHEC	***suk'k'-**	**-**	'to spin'
Had	**-**	**suk'k'-**	'to spin'
Kam	**-**	**suk'k'-**	'spin (cotton)'
Sid	**-**	**suk'k'-**	'spin (cotton)'
Ged	**-**	**suk'-**	'spin (cotton)'

The PHEC [39] stem shows pulmonic **k**, which fuses with the following **k** and glottal stop to yield **k'k'** in Hadiyya, Kambaata, and Sidaama. Gedeo shows singleton **k'**, likely reflecting degemination. Burji *sukk-* 'tie tightly, knead' ends in a geminate but without glottalization, consistent with its non-participation in fusion.

A puzzling example is [46], where glottalized **k'** appears across HEC despite the absence of an obvious conditioning environment:

PHEC [46] **ik'k'-*, **ikk-* → 'tooth' → Burji *irk'a*; PNHEC **ink'-*, **ink-* → Had *ink'e*, Kam *ink'u-ta*, and Sid *hinko*.

Sidaama *hinko* shows the expected non-glottalized **k**, consistent with broader East Cushitic cognates. The epenthetic consonants (**-n-** in NHEC, **-r-** in Burji) are irregular; possible explanations are discussed in §3.4.3 (Epenthetic -l-). Given the variation, two proto-forms are proposed at both the PHEC and PNHEC levels.

Two phonotactic factors explain the patterning of fused velars: (a) degemination in clusters – when the input **k** occurs in a consonant cluster, fusion yields singleton **k'** rather than **k'k'**, avoiding impermissible triconsonantal sequences, and (b) scarcity of stem-final **k** – non-initial ***k** is subject to lenition (§3.2.2) and elision (§3.2.1) pos-tvocalically, so fusable stem-final velars are rare except in protected environments: **Ck** clusters, **kk** geminates, and remnants of **-k** suffixes (§4.2.3).

3.3.1.4 Stem-final and Derivational ʃ, s

Stem-final sibilants **ʃ**, **s**, and **z** also undergo fusion with the glottal stop, yielding the palatal ejective **c'** (typically geminated as **c'c'**):

ʃ, **s**, **z** + **ʔ** → **c'(c')**

This process is less well-attested than labial and alveolar fusion, and some proposed reconstructions remain speculative. The label "provisional" reflects the need for additional comparative data.

[162]	**ʃ**	**Glottalized c'(c')**	**Gloss**
PNHEC	**-**	***fi-n-c'-**	'to scatter'
Had	**-**	**finc'**	'to scatter'
Kam	**fiʃʃ aʔ-**	**finc'-**	'scatter, sprinkle'
Sid	**-**	**finc'-**	'scatter, urinate'

Example [162] likely reflects the fusion of an earlier sibilant with **ʔ**. The epenthetic **-n-** is discussed in §3.4.2. This is a complicated case, and to avoid excessive abstraction, PNHEC **fi-n-c'-* is provisionally reconstructed. It is assumed that the fusion process has occurred at the PNHEC level via *fiʃ-ʔ-* and **fil-ʃ-ʔ-*. The epenthetic **-n-** seems to have replaced an older fused **l**, evident in the geminated **ʃʃ** of the Kambaata variant and found in the rest of EC.

[807]	**ʃ**	**Glottalized c'(c')**	**Gloss**
PHEC	***anʃ-**	**-**	'to chew, twist'
Burji	**anniʃ-**	**-**	'to chew'
PNHEC	**-**	***inc'-**	'to chew, twist'
Had	**-**	**inc'iic'-**	'twist, spin (cotton)'
Kam	**-**	**inc'iic'c'-**	'twist, bend'
Sid	**-**	**hiinc'i**	'chew'
Ged	**-**	**inc'a**	'chew'

Again, the fusion occurred at the PNHEC via **inʃ-ʔ*. Hadiyya has *iic'e* 'chew', which aligns semantically better with the rest of the HEC cognates, except for Kambaata.

Further examples include PHEC [221] **c'inc'-* 'good smell', [377] **c'anc'-* 'to make noise', and PHEC [671] **wad-* 'to shout' → Gedeo *woc'c'a* 'shout, make noise'.

Kambaata provides clear synchronic evidence for **z, s + ʔ → c'c'** in reflexive derivation:

> *saz-* (tr) 'advice' → *sac'c'-* (refl) 'advise for oneself' (Treis, p.c.),
>
> *t'aaz-* (tr) 'gather, assemble' → *t'aac'c'-* (refl) 'gather, come together',
>
> *has-* 'seek, want' → *hac'-c'-* 'seek, want for one's benefit' (Treis, 2022, p. 17)

This confirms the origin of **z** as a fricative when compared to the corresponding stop **d**, which would have yielded **t't'**. These synchronically transparent alternations support reconstructing the fusion rule for PNHEC.

This process appears in several derivational contexts: causative, reflexive, and denominative formations, borrowed forms, and complex morphological derivations. The development shows patterns across the HEC languages, though some forms present alternative analytical possibilities.

As with other fusion processes, Burji does not exhibit sibilant fusion (**ʃ**, **s + ʔ → c'c'**). Where glottalized consonants appear in Burji in comparable morphological contexts, they reflect either: (a) inherited non-pulmonics retained unchanged, (b) regressive assimilation of non-pulmonic consonants

to a following obstruent,[121] or (c) loanwords with source-language non-pulmonics. The absence of productive sibilant fusion in Burji further supports the classification of fusion as a shared PNHEC innovation.

3.3.1.5 Stem-final -r

In Sidaama and Gedeo, suffixal **ʔ** shows a recurrent alternation between **ʔr** and **d'**. Kawachi notes for Sidaama that "'r [ʔr] also almost always occurs as an allomorph of the middle suffix *-d* [d']" (2007, p. 33), and comparative descriptions report an alternation between implosive **d'** and a "preglottalized flap 'r'" in Sidamo, with parallel behaviour noted for Gedeo (Dawit Tilahun, 2019, p. 35, citing earlier sources). A plausible contact-induced reinterpretation (cf. Oromo influence) is that, subsequent to sonorant+**ʔ** metathesis, **ʔr** fused and was resegmented as a glottalized coronal, yielding **d'd'** and frequently surfacing as a single **d'**. This derived segment can then be generalized as an allomorph of the verbal suffix **-ʔ** (see §§4.2.3.1, 4.2.3.4).

***r** + ***ʔ** → glottalized **d'(d')**

This outcome differs from the regular fusion pattern described above for obstruents (**C + ʔ → C'C'**). It is restricted to a specific boundary sequence (**ʔr**) and to a limited set of languages, and it therefore behaves like a comparatively recent development rather than a PNHEC morphophonological process.

The glottalized **d'** attested in Gedeo, Sidaama, and in a borrowed parallel in Burji (including the reflexive allomorph **-d'**) is discussed separately in §4.2.3.4, where contact with Oromo is proposed as a conditioning factor.

[223]	**r / ø**	**Glottalized d'(d')**	**Gloss**
PHEC	***eg-**	**-**	'to wait'
Burji	**eeg-**	**-**	'to wait; watch'
PNHEC	***eger-**	**-**	'to wait'
Had	**eger-**	**-**	'wait (for)'
Kam	**eger-**	**-**	'wait (for)'
Sid	**agar**	**agad'(d')-**	'wait, watch', 'watch oneself'
Ged	**egert-**	**eged'-**	'wait!; wait (for)'

[121] Burji glottalizes non-pulmonics via assimilation to the following consonants: **t'**, **c'**, **k'** + **-t** (2sg, 3sg.f, 2pl) → **c'c'**; **t'**, **k'** + **-t-ooni** (3sg.f jussive) → **t't'**; **c'** + **-t-ooni** (3sg.f jussive) → **c'c'**.

In this set, the Sidaama and Gedeo reflexes show a split between a form with **r** and a form with **d’**, consistent with the alternation being tied to the suffixal **ʔ** environment rather than to the inherited root consonant. The same distribution also matches the observed instability of quantity in this process (**d’d’ ~ d’**).

[708], [362]	**r / ø**	**Glottalized d’(d’)**	**Gloss**
PHEC	***moor ʔ**	**-**	‘to steal’
Burji	**moorga, moorjoo**	**-**	‘thief’
PNHEC	***moor ʔ**	**-**	‘to steal’
Had	**maraat’o**	**-**	‘thief’
Kam	**mogga ʔ-**	**-**	‘steal’
Sid	**moor-**	**mood’-**	‘steal’
Ged	**moor-**	**mood’aco, moo ʔ ral-co**	‘steal’, ‘thief’

These alternations are compatible with a contact-supported resegmentation at the **ʔr** boundary, which would allow a glottalized coronal to spread as an allomorphic reflex once introduced. Based on the available evidence, the development remains restricted and does not motivate reconstructing **d’** to PNHEC as an inherited non-pulmonic phoneme.

[476]	**r / ø**	**Glottalized d’(d’)**	**Gloss**
PHEC	***c’ir-**	**-**	‘to clear’
Burji	**c’ir-**	**-**	‘to clear’
PNHEC	***c’ir-**	**-**	‘to clear’
Kam	**c’ir-**	**-**	‘pluck, tug, pull out’
Sid	**c’irati, c’ira**	**c’id’a**	‘clear one’s throat’, ‘to pluck up’, ‘to pluck at one’s hair’
Ged	**c’a ʔ ʔa**	**c’id’ata**	‘clear’, ‘clear one’s throat’

The emergence of **d’** as an alternative reflexive allomorph in Burji, Sidaama, and Gedeo is addressed in §4.2.3.4 and likely induced by contact with Oromo.

Taken together, these patterns support treating **ʔr ~ d’** in Sidaama and Gedeo as a secondary development affecting stem-final **-r** in contact with suffixal **ʔ**, with frequent leveling from **d’d’** to **d’**. The next section turns to sonorant + **ʔ** metathesis more broadly and shows why **ʔr** is the expected input for this reanalysis.

3.3.2 Metathesis

In the NHEC languages, the suffixal glottal stop **ʔ** (notably the reflexive and related derivational morphemes; see §§4.2.3.1, 4.2.5.1) shows a regular sonorant + **ʔ** metathesis: When **ʔ** follows a stem-final sonorant, the surface order is typically reversed, yielding **ʔ** immediately before the sonorant. In Sidaama and Gedeo, the specific sequence **ʔr** may additionally alternate with a glottalized **d'** (often with unstable quantity), a pattern discussed under fusion in §3.3.1.5.

$$\mathbf{C}_{[+son]}\mathbf{ʔ} \rightarrow \mathbf{ʔC}_{[+son]}$$

This is important for reconstruction because many surface strings of the form **ʔC**$_{[+son]}$ are best analyzed as underlyingly suffixal **ʔ** plus a phonological rule, rather than as infixal or stem-internal **ʔ**.

Hudson (1976, p. 250) states that metathesis is common to all the HEC languages except Burji. Such a process is attested in Sidaama (Anbessa, 2000, p. 9), Alaaba (Schneider-Blum, 2007, p. 35), Kambaata (Treis, 2008, p. 63), Hadiyya (Tadesse Sibamo, 2015, p. 39), and Gedeo (Eyob Kelemework, 2015, p. 78).

From a comparative perspective, the restriction of this metathesis to a specific phonological environment (stem-final sonorant + suffixal **ʔ**) makes it a strong candidate for a shared innovation within NHEC (and thus a potential subgrouping diagnostic in Ch. 5), especially given that Burji lacks this particular sonorant + **ʔ** metathesis. At the same time, the rule strengthens the analysis of **ʔ** as an autonomous phoneme and morpheme in PHEC/PNHEC reconstruction.[122]

Had	**r+ʔ → ʔr**	*mur-* 'cut' *muʔl-* 'cut for oneself'
Kam	**r+ʔ → ʔr**	*buur-* 'anoint/smear' → *buuʔr-* 'anoint oneself'
Sid	**l+ʔ → ʔl**	*kul-* 'tell' → *kuʔl-* 'tell for oneself'
Ged	**n+ʔ→ ʔn**	*gan-* 'hit' → *gaʔn-* 'hit onself'

Unlike NHEC languages, Burji lacks such sonorant-glottal stop metathesis, e.g., *dan-* 'able' → *danʔ-* (Tesfaye Baye, 2015, p. 183) (compare also Burji [145] *ʃinʔaa* against Gedeo *ʃiʔna* 'urine'), but has unique metathesis of the derivational (causative) marker **-s** with the stem-final velar (see Wedekind,

[122] Data sources: Kambaata (Treis, p.c.), Hadiyya (Tadesse Sibamo, 2015, p. 40), Sidaama (Anbessa Teferra, 2000, p. 22), Gedeo (Eyob Kelemework, 2015, p. 81).

1985, p. 134), e.g., *d'usk-* 'cause someone to drink', *d'ug-* 'drink' and *hiisk'-* 'cause to be closed', *hiik'-* 'close' (Tesfaye Baye, 2015, p. 76).

The evidence for **ʔr** in Kambaata and Hadiyya is limited in the available data. In Hadiyya, **r** in a cluster becomes **l** as shown in the example above (rule i(c)). In Gedeo and Sidaama, it yields an additional process **ʔ** + **r** → **d'** as discussed in §3.3.1.5, e.g., compare the previous Hadiyya and Kambaata examples with Sidaama *mur-* 'to cut' → *mud'-* 'to cut something for oneself' and *buur-* 'to apply anoint' → *buud'-* 'to apply anointment' (Indrias Xa'miso et al., 2007, pp. 389, 394, 79).

Grammars of Gedeo (Eyob Kelemework, 2015, p. 71) and Sidaama (Kawachi, 2007, p. 46) claim that the **d'** variant of the reflexive (middle) is the basic form, which is changed into a glottal stop **ʔ** when it occurs after stem-final sonorants/nasals. Such an interpretation proves to be an unnecessary complication, as there is a reconstructable, well-attested allomorph **ʔ** (§4.2.3.1). The restriction of reduction and metathesis to **d'** among non-pulmonics is another challenge for positing an intermediate **d'**. While sonorants can precede other non-pulmonics without reduction to **ʔ** or metathesis (e.g., **nt'**, **nk'**, **nc'**) – e.g., Gedeo *t'int'a* 'heel (of foot)' and Sidaama *hank'afa* 'to hug',[123] representing inherited non-pulmonic ***C'** – the prohibition of the **d'** variant of the middle in such sequence (see Wedekind, 1980, pp. 141–142) could be attributed to its external (contact-induced) origin. That would explain why the allomorph ***ʔ** is less restricted in distribution than the contact-induced (co-morpheme) **d'**.

Additionally, the presence of **d'** in the reflexive/middle suffix should be deducible from the systematic change of root/stem-final ***m** to **n** across the HEC languages when followed by such an alveolar, as discussed under §3.1.1.4. This partial assimilation is conditioned by the presence of an alveolar (**t**, **d**, **t'**, or **c'**), so **d'** should be among them. Therefore, the distribution of alveolar implosive **d'** versus the glottal stop **ʔ** variants of the reflexive can be determined by examining assimilation patterns adjacent to the nasal **m**. The lack of the expected partial stem-final **m** → **n** assimilation suggests the **ʔ** variant rather than a reduced **d'** yields **m** → **ʔm** in Gedeo and Sidaama (§3.1.1.4).[124]

[123] For an exhaustive list, see Gedeo (Eyob Kelemework, 2015, p. 196) and Sidaama (Kawachi, 2007, p. 32).
[124] Data sources: Sidaama (Kawachi, 2007, pp. 45, 33), Gedeo (Eyob Kelemework, 2015, pp. 54–55).

Sid	**m+ʔ → ʔm**	*waam-* ‘to bury’ → *waaʔm-*
	m+ʔ → ʔm	*tum-* ‘to mix (honey) with water’ → *tuʔm-*
Ged	**m+ʔ → ʔm**	*tum-* ‘pound’ → *tuʔm-a* ‘to pound for oneself’
	m+ʔ→ ʔm	*harum-* ‘short’ → *haruʔm-* ‘become short’

There is no evidence that the metathesized suffix was **d’** in the above examples since none of them yield **ʔn**. Hence, the allomorph **ʔ** is the one attached, and there is no need to posit an intermediate **d’**.

Two examples with **ʔn**, corresponding to **ʔm**, look like they may challenge the lack of partial assimilation argued above:

[75]	**ø/ʔ**	**ʔm**	**ʔn**	**Gloss**
PNHEC	***t’eʔm-**	**-**	**-**	‘to taste good’
Had	**t’eeʔ-**	**-**	**t’eʔna**	‘be sweet, taste good’, ‘taste, flavor’
Kam	**t’eʔ-**	**-**	**t’eeʔna**	‘taste good, become sweet’, ‘taste, flavor’
Sid	**-**	**t’uʔma**	**-**	‘to be good’
Ged	**-**	**t’eʔme**	**-**	‘to be good’

In [75], the lack of **m → n** assimilation in Gedeo and Sidaama suggests **ʔ** in the stem **t’eʔm-*. However, Kambaata *t’eeʔna* involves a nominal derivation with **n** (Treis, p.c.) from the verb *t’eʔ-* ‘taste good, be(come) sweet’, which applies to the Hadiyya word as well since the verb is *t’eeʔ-* ‘to be expensive, be sweet, taste good’ (see Burji *d’ed’emm-* and Oromo *d’amd’ama*). Therefore, there is still no need to posit an intermediate **d’**.

[170]	**ø/ʔ**	**ʔn**	**Gloss**
PHEC	***sum ʔ**		‘urinate’
Burji		**sinʔ-aaw-**	‘urinate’
PNHEC	***ʃum ʔ**		‘urinate’
Had	**ʃum-eʔ-**		‘urinate’
Kam	**ʃum-aʔ-**		‘urinate’
Sid	**ʃumi, ʃumaa-ʔr-**		‘urinate’
Ged		**ʃiʔn-**	‘urinate’

In set [170], the **(V)ʔ** sequences are reflexes of the support verb **ʔ* ‘to do’ (§4.2.4.2), with an **a** inserted before it. However, Sidaama **ʔr** in *ʃumaa-ʔr-* is an allophone of the reflexive marker **d’** (§4.2.3.1), whereas the Sidaama reflex of the support verb **ʔ* ‘to do’ appears as an allomorph **aa** that then cliticizes to the stem. In addition to the irregular **n**, the irregular stem vowel **i** in Burji and Gedeo (vs. PNHEC **ʃum ʔ* ‘urinate’) motivates checking for contact

effects before treating the stem shape as inherited (cf. Oromo *finc'-* → Sidaama *finc'i* 'urinate'). It is likely that Gedeo and Burji words with **n** have been influenced by the Oromo parallel. Thus, the PNHEC word is **ʃum ʔ* 'urinate', reconstructed with **m** (see also PHEC **sum-* [145] 'urine').[125]

3.3.3 Elision

Across HEC, non-initial **ʔ** is frequently unstable, and its loss is often morphophonologically conditioned. In Burji, Sasse (1982, pp. 18, 112) describes compensatory vowel doubling when **ʔ** is lost before a consonant (e.g., a preconsonantal stem alternation such as *kaʔ-* vs *kaa-* 'get up'). Comparable deletion with compensatory doubling is also reported for Gedeo (Eyob Kelemework, 2015, p. 84), and Wedekind (1980, p. 132) notes striking elision of **ʔ** (and /**h**/) in Sidaama. These patterns mean that the absence of surface **ʔ** in the expected environments, vowel doubling, and other alternations can be the primary comparative cues for earlier ***ʔ**.

In Burji, patterns beyond compensatory doubling include (a) assimilation of **ʔ** (or *y*) + **s** → **ss**, and (b) correspondences where Burji **y** matches NHEC **ʔ** in comparable morphological contexts. The first pattern is straightforwardly an alternation conditioned by full assimilation of stem-final and **s**; the second requires careful analysis, because a Burji **y** correspondence could reflect multiple diachronic sources (including the glide ***y**, epenthesis, or morphologized allomorphy). The key comparative point is that all of these alternations can systematically mask underlying **ʔ**, and therefore must be controlled for when comparing reflexive/denominative morphology across HEC.

[125] Metathesis is not restricted to the glottal stop **ʔ**. In Gedeo, Eyob Kelemework (2015, pp. 79–80) lists the following additional kinds of stem-final obstruent and suffixed sonorant metatheses: stem-final stops/affricates/fricatives and a suffix-initial **n** and stem-final obstruents and suffix-initial **l**. Similarly, in Hadiyya, metathesis of the root-final obstruent, such as **t** and nasal **n**, e.g., *ʔit-* 'eat' *ʔint-úmmo* 'we ate' (Anbessa Teferra, 2014, p. 39).

[269]	**ʔ**	**y**	**Assimilated**	**Gloss**
PHEC	***t'aʔ-, *t'eʔ-**	**-**	**-**	'to extinguish, to diminish'
Burji	**-**	**t'ay-, t'ey-**	**t'ass-**	'extinguish, dry out' (intr), 'extinguish, dry up' (tr)
PNHEC	***t'aʔ-, *t'eʔ-**	**-**	**-**	'to extinguish, to diminish'
Had	**t'aʔ-**	**-**	**-**	'be depleted, diminish'
Kam	**t'aʔu, t'eʔ-**	**-**	**-**	'extinguish, be depleted'
Sid	**t'eʔ-**	**-**	**-**	'be depleted, be insufficient'
Ged	**t'eʔy-**	**-**	**-**	'be incomplete'

Both the NHEC **ʔ** and Burji (and Gedeo) **y** are likely derivational suffixes (§4.2.5) or support verbs (§4.2.4) attached to a root that ends in a glottal stop, as evident in Gedeo with both **ʔ** and **y**. In Burji, the meaning of the intransitive *t'ay-* is 'to be dried, be emptied' according to Getahun Addo (2017, p. 823). The second variant *t'ass-* is transitive and segmented as *t'as-s-* 'extinguish, dry up, drain' in Lamberti & Sottile (1997, p. 539). The geminate **ss** is the derivational **s** assimilated to the stem-final **ʔ** or derivational **y** (§3.1.2.2).

[34]	**ʔ**	**y**	**Assim./Lost**	**Gloss**
PHEC	***baʔ-**	**-**	**-**	'be lost'
Burji	**-**	**bay-, bey-**	**ba-ss, be-**	'be lost', 'get lost'
PNHEC	***baʔ-**	**-**	**-**	'be lost'
Had	**biʔ-**	**-**	**-**	'to be lost'
Kam	**baʔ-**	**-**	**-**	'be lost'
Sid	**baʔi**	**-**	**-**	'perish, be destroyed'
Ged	**baʔ-**	**-**	**-**	'be lost'

The causative in Burji is *ba-ss,* and there is a variant without the epenthetic **-y-**, i.e., *be-* 'get lost' (Sasse, 1982, p. 34).

[749]	**ʔ**	**Lost**	**Gloss**
PHEC	***yaaʔ-**	**-**	‘assembly, meeting’
Burji	**-**	**yaa**	‘assembly, meeting’
PNHEC	***yaaʔ-**	**-**	‘assembly, meeting’
Had	**yaaʔa**	**-**	‘to call, to be gathered’
Kam	**yaaʔa**	**-**	‘assembly of elders’
Sid	**yaaʔa**	**-**	‘communal assembly that holds reconciliation of bloodshed’
Ged	**yaaʔa**	**-**	‘crowd, moving crowd’

In Burji, the derived transitive verb *yaaʔis-* ‘to come together and be accumulated and assembled’ (Getahun Addo, 2017, p. 823) is separated from the glottal stop with **i**, which is likely a reflex of the derivational **y** (§3.1.2.2). The intransitive verb is *yaas-* ‘to go or move in procession line without discontinuation’, without the glottal stop or evidence of fusion, so it must have been dropped.

[117]	**ʔ**	**y**	**Lost**	**Gloss**
PHEC	***saʔ-**	**-**	**-**	‘cow’
Burji	**-**	**saa-yi**	**saa**	‘cow’ (subj), ‘cow’
PNHEC	***saʔ-**	**-**	**-**	‘cow’
Had	**-**	**saáyyạ**	**-**	‘cow’
Kam	**saʔa**	**-**	**-**	‘cow’
Sid	**-**	**-**	**saa**	‘cow’
Ged	**-**	**saayya**	**-**	‘cow’

The **-y-** in Burji, Hadiyya, and Gedeo is epenthetic. Therefore, it only occurs intervocalically in the subject in Burji *saa-yi* (subj) and plural *saa-yanna* (pl) (Sasse, 1982, p. 161).

Additional sets showing ***ʔ** elision include PHEC [440] **wee?n-* ‘Colobus monkey’, PHEC [168] **hee?-* ‘to exist, to live’ → Burji *heettaa* (interj) ‘Here it is!’, PHEC [30] **ka?-* ‘to get up, arise’ → Burji *ka??-* (Roba Dame & Wedekind, 2008, p. 49), *ka?-* (Sasse, 1982, p. 112), *kaa-* (preconsonantal) ‘get up, arise’ (Sasse, 1982, p. 112), Sid *ka-i-s-* (agentive) ‘to lift up’ (Kawachi, 2007, p. 734), and [114] PHEC **debe?-* ‘to belch’ → Burji *demmees-* and *deemeeys-*.[126]

[126] The change from PHEC [114] **debe?-* ‘to belch’ to Burji *demmees-* and *deemmeys-* involves multiple stages: ***-n-** epenthesis produces **denbe?-*, assimilation of ***nb** produces **demme?-* (§3.4.3), then upon the addition of the derivational **-s** it became

Loss, assimilation, and epenthesis are evident in the Burji examples.[127] Some of the morphemes that trigger the insertion, assimilation, and loss include the subject or nominative suffix **-i** in Sidaama [117] and causative **-s** in Burji [34]. Historically, other morphemes might have had the same impact on the non-initial **ʔ**.

The section on the variation in the stem vowel in 2.2.2 provides further analysis of most of the above words. The epenthetic function of **ʔ** in the HEC languages is discussed in the next Section 3.3.4.

The loss of non-initial **ʔ** and similar outcomes may have arisen independently in multiple languages. The elision interacts with a specific morphological marker (e.g., the reflexive/denominative **ʔ** treated in §4.2.3.1 and §4.2.5.1, or the support verb ***ʔ** 'to do' in §4.2.4.2), where the conditioning environment and the type of compensatory effects provide structured evidence for the underlying ***ʔ**.

3.3.4 Epenthesis

Beyond the inherited segmental system, the HEC languages apply morphophonological repairs when morphology creates impermissible vowel sequences. This section examines intervocalic **-ʔ-** insertion as a hiatus-resolution strategy, most consistently attested in Hadiyya and Gedeo, and compares it with glide insertion and vowel coalescence in the remaining HEC languages. The discussion treats **-ʔ-** primarily as a repair operating across morpheme boundaries (especially in vowel-initial suffixation to vowel-final stems[128]), though some cases may reflect earlier consonant loss and later reanalysis. Distinguishing inherited segments from later epenthesis is necessary for reconstructing stem shapes and for separating shared innovations from parallel developments.

---

demmee-s- and upon the addition of the support verb **y** it became *deemeeys* (§3.4.1.2), with puzzling V_1 doubling.

[127] Hadiyya and Gedeo have *saayya* 'cow', but it might include a suffix (§4.1.1.2 (Collective *-ey-e (f))). A few exceptions in Sidaama include [30] *ka-i-s-* (agentive) 'to lift up' and [117] *saa* 'cow'.

[128] The term "vowel-final stem" is used here to cover both terminal vowels (§2.2.3) and stems ending in a vowel but which potentially have an underlying consonant.

[382]	P(N)HEC	-ʔ-	VV, V́, V	Gloss
PHEC	***oll-**	-	-	‘neighbour’
Burji	-	-	**olla, ollá**	‘neighbour’
PNHEC	***oll-**	-	-	‘neighbour’
Had	-	**ollaʔa**	-	‘neighbour’
Kam	-	-	**olloo**	‘neighbour’
Sid	-	-	**ollaa**	‘neighbour’
Ged	-	**ollaʔa**	-	‘neighbour’

Burji has *óll-a*, *oll-ée* ‘village’ as well. Sasse (1982, p. 155) considers [382] to be from Oromo *ollaa* ‘neighbor’.

[212]	P(N)HEC	-ʔ-	VV, V́, V	Gloss
PNHEC	***ar-**	-	-	‘husband’
Had	-	**aroʔo**	-	‘husband’
Kam	-	-	**aroo**	‘husband’
Sid	-	-	**ʔaro, aroo, aró**	‘husband’
Ged	-	**aroʔo**	-	‘husband’

The additional **o** is the case marker distinguishing PNHEC [212] **ar-o* (m) ‘husband’ from [271] **ar-e* (f) ‘wife’.

[350]	P(N)HEC	-ʔ-	VV, V́, V	Gloss
PHEC	***sang-**	-	-	‘ox, bull’
Burji	-	-	**sangaa**	‘bull’
PNHEC	***sang-**	-	-	‘ox, bull’
Had	-	**sangaʔa**	-	‘ox for slaughter’
Kam	-	-	**sanga**	‘ox (castrated)’
Sid	-	-	**sanga**	‘stallion, bullock’
Ged	-	**sangaʔa**	**sanga**	‘ox’, ‘ox for beef’

Set [350] has parallels in Amharic *sänga* and Oromo *sangaa* ‘ox for beef’.

Other examples with epenthetic **-ʔ-** in Hadiyya only include: PHEC [478] **kott-* ‘hoof, foot’ → Had *konteʔe* ‘hoof, foot, foot print’ and [91] **mooyy-* ‘mortar’ → Had *mooyyeʔe*.[129]

Variations in transcription, particularly vowel quantity and quality, exist across different sources for the above examples: Burji [382] *ollaa* (Roba Dame & Wedekind, 2008, p. 67), *óll-a*, *oll-ée* (Sasse, 1982, p. 154), *ollá* (Hudson,

[129] These also have parallels in Oromo: [478] *kottee* ‘hoof’ and [91] *mooyyee* ‘mortar’.

1989, p. 104) ‘neighbor’; [350] *sangaa* ‘bull’ (Roba Dame & Wedekind, 2008, p. 70), *sanga* ‘ox for beef’ (Hudson, 1989, p. 109); [478] *kottee* ‘hoof’ (Roba Dame & Wedekind, 2008, p. 53), *kotte* ‘track’ (Hudson, 1989, p. 81); [91] *mooyyee* (Roba Dame & Wedekind, 2008, p. 63), *moyyé*, *mooyee* (Hudson, 1989, p. 101) ‘mortar’. The same applies to Sidaama [212] *ʔaro* (Yri & Pepper, 2019), *aroo* (Indrias Xa’miso et al., 2007, p. 31), *aroo*, *aró* (Hudson, 1989, p. 82) ‘husband’, Gedeo [350] *sangaʔa* ‘ox’ (Roba Dame & Wedekind, 2008, p. 8), *sanga* ‘ox for beef’ (Hudson, 1989, p. 109). As evident in some of the corresponding vowels in the languages without epenthesis, the vowel preceding the inserted glottal stop was either double or accented originally. Therefore, the variation in the transcription of the vowels potentially reflects dialectal differences or transcription choices.

The glottal stop ***ʔ** can only be confidently reconstructed in PHEC for monosyllabic stems (§2.1.5.2 (PHEC *ʔ)). Therefore, instances of intervocalic **-ʔ-** in potentially disyllabic or multisyllabic reconstructed stems might be attributable to later, language-specific epenthesis (e.g., PNHEC or later stages), rather than direct inheritance from PHEC, unless they can be shown to arise from earlier suffixation across a morpheme boundary.[130]

3.3.4.1 Hiatus Repair -ʔ-

The most common environment for epenthetic **-ʔ-** is intervocalically, arising when a vowel-initial suffix is added to a vowel-final stem (**V-ʔ-V**). This process is particularly active in the nominal and pronominal morphology of Hadiyya and Gedeo.

The extended epenthesis environment can be further specified because it has been attested intervocalically between two to three (even four) vowels with different accentuation. Additionally, there are instances of an underlying consonant surfacing and reclaiming the epenthesis slot. Those three extended epenthesis environments are discussed in the following sections.

Two Vowels (V+V)

Intervocalic epenthesis between two vowels has already been demonstrated in all the examples above ([382], [212], and [350]). Below are additional

[130] Tadesse Sibamo (2015, p. 59) reports glottal stop epenthesis between the final vowel of disyllabic proper nouns but not trisyllabic nouns in Hadiyya.

examples based on synchronic morphological triggers, i.e., vowel-initial suffixes.

		Singular	**Plurative/Collective**	**Gloss**
Ged	[299]	**daanalco**	**daanaʔe (pl.m), daanaʔo (coll.m)**[131]	‘elephant’
Had		**daaneecco**	**daaneʔe**	‘elephant’
Had	[628]	**anna**	**annoʔo**	‘father’
Had	[547]	**ama, umma**	**ammoʔo**	‘mother’
Had	[150]	**baalleʔe**	**baalleʔuwwa**	‘feather’

The examples show that plurative or other suffixes trigger **-ʔ-** insertion in Hadiyya and Gedeo (e.g., [628] Hadiyya *anna* ‘father’, pl: *annoʔo*; Gedeo *daanalco* ‘elephant’, pl: *daanaʔe*) compared to vowel changes or **-w-** epenthesis (§3.4.1.2) in cognates (e.g., [299] Sidaama *daaniwo* ‘elephants’). This highlights **-ʔ-** as a language-specific strategy for hiatus resolution in these contexts.[132]

Here are the cognates of the above examples in the rest of HEC:

> Kam [299] *zaneeccu*, *zanooccu*, Sid *daaniicco* (sg), *daaniwo* (pl) ‘elephant’.
>
> Burji [628] *anaa* ‘man, husband’, Kam *anna* ‘father, owner, parent’, Sid *ʔanna* ‘father, owner’, Ged *anna* ‘father’.
>
> Burji [547] *ama* ‘woman (wife)’, *amá* ‘mother’, Kam *ama-ta* ‘mother’, Sid *ʔama* ‘mother’, Ged *ama* ‘mother’.
>
> Burji [150] *baallee*, *baalle*, *baalée*, Kam *baalli-ta*, Sid *baallicco*, *balle* (see Oro *baallee* ‘feather’, Wollega Oromo *baal-eʔ* ‘feather’).

The above examples prove that **-ʔ-** epenthesis is triggered by vowel-initial suffixation. It has also been observed that either the stem-final (terminal) vowel assimilates to the vowel of the suffix or the terminal vowel copies the quality of the preceding vowel.

Hayward (1997, pp. 111–112) proposed vowel doubling and **-ʔ-** epenthesis in Hadiyya, Kambaata, and Gedeo. He describes it as a process of

---

[131] See *Gedeʔo* (coll.m) ‘Gedeo people’ (Wedekind, 2008, p. 88)
[132] See also the accusative Gedeo *amaaʔa* (Eyob Kelemework, 2015, p. 245) of [547] *ama* ‘mother’ and *godobeeʔe* (acc) of *godoba* ‘stomach’.

resyllabification of the two vowel morae, which are originally long terminal vowels, into two separate syllables. In this process, the inserted glottal stop represents an onset for the second syllable, in the form of what he calls “minimal consonantality”. In Hadiyya, this process has affected long terminal vowels with any of the three qualities available, i.e., **aa**, **oo**, and **ee**, e.g.:

> *aroʔo* ‘husband’, *buraʔa* ‘straw’, *k’ureʔe* ‘jar’.

However, in Kambaata, the process operates in the case of earlier ***aa**, but does not appear to have affected ***ee** or ***oo**, which sometimes remain as double vowels, e.g.:

> *buraʔa* ‘straw’ vs. *aroo* ‘husband’, *eloo* ‘well’, and *zanoo*, *zanee(-ccu)* ‘elephant’.

Gedeo, on the other hand, displays features intermediate to the preceding languages, splitting earlier long ***aa** and ***oo** terminal vowels into two syllables in many words while retaining unchanged reflexes of long ***aa** and ***oo**, as well as long ***ee**, in others.

Similar analysis of Hadiyya is found in Tadesse Sibamo (2015, p. 59), where he posits that “[d]isyllabic proper nouns (both masculine and feminine) duplicate their final vowel and insert the glottal stop to avoid vowel cluster in absolutive... The duplicated vowel of the masculine proper nouns is replaced by nominative case suffix… -i”, e.g.:

> *deéro* → *deeróʔo̩* (Absolutive), *deeróʔo* (f.nom) and *lobe* → *lobéʔe̩* (absolutive) *lobéʔ-i* (nom.m),
>
> *buubaáme* → *buubaáme̩* (absolutive), *buubaáme* (f.nom) (trisyllabic)

Hayward’s and Tadesse Sibamo’s insertion analysis is accepted here. However, the assumption of vowel duplication is a less parsimonious explanation than the proposal of epenthesis and suffixation (e.g., nominative **-i** in Hadiyya). Additionally, the vowels in the environment of the epenthesis vary in quality, quantity, and accentuation, which means the vowels around epenthetic **-ʔ-** are not always identical. Therefore, my analysis posits that the epenthesis operates within morphophonological processes triggered by

vowel-initial suffixes to explain the various qualities, quantities, and accentuation of the corresponding vowels, e.g.:[133]

	-ʔ-	**VV**	**V́, V**	**Gloss**
Burji	**-**	**támboo**	**tambó**	‘tobacco’
Had	**tambaʔa**	**-**	**-**	‘tobacco’
Kam	**tumbeʔo**	**-**	**-**	‘tobacco’
Sid	**-**	**-**	**tambo**	‘tobacco’
Ged	**-**	**-**	**tambo**	‘tobacco’

See Oromo *tamboo*, *timboo* ‘tobacco’.

	-ʔ-	**VV**	**Gloss**
Had	**antabaʔa, antabaʔ-icco**	**-**	‘hen’
Kam	**antabeeʔu**	**antabee-ccu-ta**	‘hen’

The examples show variation in the quality and quantity of the vowels preceding and following the epenthetic **-ʔ-**. For instance, the first example might be an Amharic loan *təmbaho* ‘tobacco’ (Leslau, 1976, p. 100). Therefore, upon the elision of the intervocalic **-h-** (§3.2.2), the resulting **a**-**o** vowel sequence coalesced into **-oo** in Burji, Sidaama, and Gedeo, while Hadiyya and Kambaata inserted epenthetic **-ʔ-**. The surrounding vowels show further assimilation effects in Hadiyya (**aʔa**) but none in Kambaata (**eʔo**). Ritter (2007, p. 9) transcribes the example from Hadiyya as follows: *antabaaʔa* (pl), *antabaakkicco* ‘chicken, hen’. The surfacing of **kk** in the singulative is not an isolated case and is addressed below.

Sidaama has only a marginal instance of glottal epenthesis between the negative prefix and the verb stem *ag-* ‘drink’, for example, in *di-agoomm-o* → *diʔagoommo* ‘I have not drunk’ (Anbessa Teferra, 2000, p. 28). This is comparable to a rather optional epenthesis in Gedeo, where Eyob Kelemework (2015, p. 324) notes that speakers may use any of the variant forms *joʔokini*, *jookin*, *joʔokan*, or *jookan* ‘or’, so the glottal stop is either deleted or not inserted.

The above epenthesis was arguably originally to separate the accented and unaccented vowel **V́$_{TV}$-ʔ-V** (shifting from ***V$_{TV}$-V́**) sequence upon the addition of a vowel-initial suffix (§2.4).

[133] Data source: Hudson (1989, pp. 156, 39).

Three Vowels (V+VV, VV+V)

Epenthesis resolves more complex vowel hiatus (**VV-ʔ-V** or **V-ʔ-VV**), primarily documented in Gedeo and Kambaata pronominals and certain derived nominal forms, resolving potential triphthongs or sequences across morpheme boundaries created by suffixation.[134]

	Nominative	**Accusative**		**Gloss**
Kam	**naʔóo-t**	**-**	(1pl)	‘we’
	aʔnoʔóo-t	**kiʔneʔée-ta**	(2pl)	‘you (pl)’
	issoʔóo-t	**issoʔóo-ta**	(3pl)	‘they’
Ged	**noʔo**	**noʔo-oʔo**	(1pl)	‘we’
	haʔnno	**haʔnno-oʔo**	(2pl)	‘you (pl)’
	insaʔnne	**insaʔnne-eʔe**	(3pl)	‘they’
	itti	**ettene-eʔe**	(f)	‘that (f)’
	ikki	**okkone-eʔe**	(m)	‘that (m)’
Sid	**tiiʔi**	**teʔe, teʔenne**	(f)	‘over there’
	kuʔu	**koʔo, koʔonne**	(m)	‘over there’

The examples illustrate **-ʔ-** insertion within complex pronominal forms in Kambaata (e.g., *naʔóo-t* (1pl.nom) ‘we’) and Gedeo (e.g., *noʔo-oʔo* (1pl.acc) ‘we’), and potentially Sidaama demonstratives. The precise placement relative to accent and vowel quantity differs: **V-ʔV́V** in Kambaata vs. **V́-VʔV** in Gedeo plurals. However, the function of resolving vowel sequences introduced by suffixes (likely case/number markers) appears similar.

The **-ʔ-** epenthesis is common between 3–4 vowels and between accented and unaccented vowels in Kambaata and is not confined to the pronominals. Treis (2008, p. 189) identifies such cases of epenthetic glottals occurring before the accented/double vowels of the genitive and ablative forms as an unknown potential extension, e.g.:

> *ber-é* (acc) ‘yesterday’, *ber-é ~ ber-e-ʔée* (gen), *ber-éecc ~ ber-e-ʔéecc* (ablative) (compare with the dative *ber-ée*),
>
> *t’on-íta* (acc) ‘before, recently’, *t’on-é ~ t’on-e-ʔée* (gen).

Later, Treis identifies that glottal extension (at least for the genitive) as a remnant of the associative morpheme. Then, another comparable glottal stop

[134] Data sources: Gedeo (Eyob Kelemework, 2015, p. 112), Kambaata (Treis, 2008, pp. 330–335), Sidaama (Kawachi, 2007, pp. 189, 200).

has been described as inserted before derivational morphemes (Treis, 2008, p. 275), e.g.:

> *ceemm-í-ta* (acc), *ceemm-é* (gen) ‘laziness’, *ceemm-eʔ-aam-ú* (m.adj), *ceemm-eʔ-aam-í-ta* (f.adj) ‘lazy’.

The latter analysis applies to most, if not all, of the other unknown instances, including the pronouns, and has comparative support.

In Gedeo, in addition to the above examples, all persons in the accusative show such a pattern of suffixation and epenthesis. Eyob Kelemework (2015, p. 112) describes the process as lengthening the final vowel with an intermittent glottal stop. He adds that, since the final citation vowels (**e** and **o**) in the accusative case carry high pitch, the following double vowel plus an intermittent glottal stop may optionally be dropped due to its unstressed nature. Thus, the above accusative forms may alternatively be *noʔó* (1pl), *haʔnnó* (2pl), and *insaʔnné* (3pl) with a high pitch on the final vowels. The same is true for demonstratives. Eyob Kelemework’s analysis assumes the suffixation of the citation vowels **-e** and **-o** in the accusative, which contradicts the marked case (accusative vs. citation) and requires what he calls a change of the citation vowel before the alleged “lengthening,” e.g., **-i** → **-e** in *itti* (f.nom), *ettene-eʔe* (f.acc) ‘that’ (f). Unless the accusative is equivalent to the citation form, it is more plausible to identify what he calls “citation vowels” as accusative vowels/suffixes or at least attribute the change of the nominative vowel to assimilation to the attached vowel. Contrary to Kambaata, such epenthesis is chiefly, if not exclusively, found in the pronominals in Gedeo.[135]

Contrary to Hayward’s (1997) analysis focusing solely on the splitting of terminal double vowels, the pronominal data in Gedeo and Kambaata appear to involve the suffixation of an additional accented double vowel (**V́V**) following the terminal vowel of the stem. Additionally, the epenthesis is between the vowels of the suffix -**V́V** in Gedeo (**V́-VʔV**), but before them in Kambaata (**V-ʔV́V**) on the pronouns. That explains the assimilation of the terminal vowel to the vowel of the suffix in Gedeo. Such epenthesis resolves the impermissible **VVV** sequence introduced by a vowel-initial suffix.

[135] Wedekind (1980, pp. 140, 161) reports that such a process is present phonetically, i.e., it could be heard in fast speech even on common nouns and characterized with a unique laryngealization of the terminal vowel in Gedeo.

It is unclear if the Sidaama demonstrative examples involve an epenthetic or an independent suffix with a glottal stop. However, the pattern aligns well with the Gedeo and Kambaata epenthesis triggered by case vowel suffixation.

Underlying Consonant

In some cases, epenthetic **-ʔ-** in Hadiyya or Gedeo appears in positions where an underlying consonant surfaces in specific morphological contexts (e.g., before the Hadiyya singular suffix **-icco**) or where cognates in other HEC languages show a consonant (often **k** or **w**). This raises the question of whether **-ʔ-** is merely filling a hiatus created by the loss of an earlier consonant, or if it is directly inserted in a different intervocalic environment which has nothing to do with consonant replacement (see the first subsection: Epenthesis Between Two Vowels).[136]

[795]	**-ʔ-**	**C**	**ø**	**Gloss**
Had	**fellaʔa**	**fellak-icco**	**-**	‘goat’ (pl), (sg)
Kam	**fellaaʔu**	**-**	**felleeccu**	‘goat’, ‘billy goat’

[273]	**-ʔ-**	**C**	**ø**	**Gloss**
Burji	**-**	**-**	**biinnee**	‘mosquito, fly’
Had	**biimbeʔe**	**biimbekicco, biimbekkícc**	**-**	‘mosquito, fly’
Kam	**-**	**-**	**biimbi-ta**	‘mosquito, fly’
Sid	**-**	**-**	**biinne**	‘mosquito, fly’
Ged	**-**	**-**	**biiʔne**	‘mosquito, fly’

[382]	**-ʔ-**	**C**	**ø**	**Gloss**
Burji	**-**	**-**	**ollaa, oll-áccaa**	‘neighbour’
Had	**ollaʔa**	**ollakkicco**	**-**	‘neighbour’
Kam	**-**	**-**	**olloo, olleeccu**	‘neighbour’, ‘villager’
Sid	**-**	**-**	**ollaa, olliicco**	‘neighbour’
Ged	**ollaʔa**	**-**	**-**	‘neighbour’

[136] The intervocalic gemination of single **k** → **kk** /**V_V** is comparable to the intervocalic glide gemination ***w** → **ww**, ***y** → **yy** /**V_V**. They share the vulnerability to elision in intervocalic position (see §3.2.2 and iii). Additionally, most of the listed examples are nominals. The following might involve a verbal suffix **-ʔ** and **-eek**: Had *fooʃ-eʔ-*, Kam *fooʃ-eeʔ-*, *fooʃ-eek-* ‘breathe’. Treis (p.c.) corrects Hudson’s transcription to *fooʃʃ-eeh-* ‘breathe’.

The Oromo parallel is *ollaa* ‘neighbor’.

[252]	**-ʔ-**	**C**	**Gloss**
Burji	**-**	**hasaaw-**	‘to talk’
Had	**heesseʔ-**	**hasaaw-**	‘to speak figuratively’, ‘chat’
Kam	**-**	**haasaaww-**	‘discuss’
Sid	**-**	**hasaaw-**	‘speak, chat, talk’
Ged	**haasoʔ**	**-**	‘speak, chat’

More examples are found in the following comparative sets:

Had [446] *boonkeʔe* ‘pig, wild pig’, *boonkekkicco* (sg) ‘pig, wild pig’: Burji *booyyee*, *booyyé*, Kam *bookki-ta*, *bookk-iccu* (sg), Ged *booyye* ‘pig, wild pig’ (see Oro *boyyee* ‘pig, wild pig’)

Had [267] *t’uraʔa* ‘claw, finger-nail’: Burji *t’unga*, Had *t’urank’a*, Kam *t’ulunk’a-ta* (Kam Geyyota), *t’ulunga-ta* (Kam Enano), Sid *c’ulunk’a* ‘claw, finger-nail’

Ged [1034] *akkaʔo* ‘grandfather’: Sid *ahaaho* ‘grandfather’, Burji *akki* ‘grandfather’ (see Oro *akaakayyu* ‘grandfather’)

Ged *lolaʔa* ‘flood, torrent’: Sid *lolahe*, *lolahá* ‘flood, torrent’, Burji *lolá* ‘flood, torrent’ (see Oro *lolaa* ‘flood, torrent’) (Hudson, 1989, p. 65)

These examples show Hadiyya and Gedeo **-ʔ-** corresponding to **-k-** in Hadiyya (before **-icco**) and to **k**/**w**/**ø** in other languages. For instance, Hadiyya *fellaʔa* (pl) vs. *fellak-icco* (sg) ‘goat’; Gedeo *akkaʔo* ‘grandfather’ vs. Burji *akki*; Gedeo *haasoʔ* ‘speak’ vs. Kambaata *haasaaww-* ‘discuss’. This pattern suggests that **-ʔ-** often occupies a slot where an underlying consonant (**k** or **w**) has been elided intervocalically.[137]

The glottal stop in examples [795], [273], [382], and [446] corresponds to a consonant in Hadiyya. The underlying consonant surfacing before the singulative suffix **-icco** is mainly the velar **k**, which is susceptible to lenition (§3.2.2) in such a post-vocalic position; therefore, it sometimes surfaces as a geminate. In sets [1034], [252], and the last example, *lolaʔa* ‘flood, torrent’, Gedeo has a glottal stop corresponding to elided **k** and **w** in other HEC languages. The geminate **nn** in Burji and Sidaama might be due to the fusion

[137] The formal shape of **ʔ** and **w** in the reflexes of the cognate set [252] suggests an alternative analysis as derivational suffixes (§4.2.5).

of **ʔn* in [273], where the glottal stop cannot be analyzed as epenthetic but rather as part of the stem. The correspondence with **ø** in languages other than Gedeo and Hadiyya involves mostly double or accented vowels, particularly in Burji.

Reflexes of the example [267] PHEC **t'ulak-* 'claw, finger-nail', including epenthetic **-ʔ-**, could be explained via morphophonological processes:

> Lenition (§3.2.2), and epenthesis: **t'uraha* → Had *t'uraʔa*
>
> Gemination (§3.1.1) and epenthesis (§3.4.2): **t'ulak~k-* → Kam *t'ulunga-ta* Burji *t'unga*
>
> Fusion (§3.3.1.3) and epenthesis (§3.4.2): **t'ulak-ʔ* → **t'ulak'k-* → Had *t'urank'a*, Kam *t'ulunk'a-ta*, Sid *c'ulunk'a*

An alternative analysis would be to posit separate processes, i.e., instead of underlying **C** replaced by **-ʔ**, both the surfacing of the consonant and the insertion of the glottal stop are independent of each other (e.g., [252]). As discussed above, the glottal stop is inserted as part of the usual epenthesis process before a vowel-initial suffix. The consonants are from different sources, including stem-final (Gedeo [267], [1034], and [252]) and a suffix with a voiceless velar (Hadiyya [795], [273], [382], and [446]), e.g., **-ga** and **-da** in Burji *duwwaaga* (m.adj), *duwwaada* (f.adj) corresponding to an epenthetic glottal stop in Hadiyya *daamaʔa* 'red'. That means the latter is added as part of the suffixation process, e.g., singulative **-k-icco** in Hadiyya. However, such a hypothesis fails to account for cases where the elided consonant is part of the root/stem, as evident in the cognates, particularly in Gedeo (See [267]).

Additionally, in Hadiyya, Tadesse Sibamo (2015, pp. 100–102) reports additional subject markers occurring as geminated **mm** (1sg/1pl), **tt** (2sg), and **kk** (3sg.m/3pl). However, the slot of that additional subject marker is occupied by an epenthetic **-ʔ-** instead of the expected **-tt-** in 2pl and 3sg. The epenthetic **-ʔ-** is inserted in that slot between the two vowels, i.e., aspect **-ó** and terminal or indicative mood vowel **-o**, e.g.,

ʔuww-	'give'	+	2pl	**ʔuww-i-t-akk-ó-ʔ-o**
		+	3sg.f	**ʔuww-i-t-ó-ʔ-o**

Tadesse Sibamo (2015, pp. 103–104) refers to a different analysis by Crass (2013), who calls the terminal vowels and the preceding consonantal

morphemes “additional morphemes”. Since these vowels are identical in all verb forms, Crass does not consider them aspect markers. For Crass, the main verbs in the HEC languages consist of a lexical base, a paradigm of subject markers (see **-t** above), an aspect marker (see **-ó** above), and a paradigm of “additional morphemes” (see **-ʔ-o**). The **-akk-** in 2pl is a plurative marker (Tadesse Sibamo, 2015, p. 100).

There are numerous further examples from Hadiyya and Kambaata (Hudson, 1989, pp. 146, 36, 43, 119, 31).

-ʔ-	**-k-/-t-/-ø-**	
Had **abbaʔa**	Kam **abba-ta**	‘Sunday’
Had **duhaʔa**	Kam **duuha-ta**	‘condition’
Kam **t’araʔa** (pl)	Had **t’arak-icco** (sg)	‘python’
Kam **gaamelaʔa, gaameela**	Had **gaamelak-icco** (sg), **gaamela**	‘camel’

The above examples raise a question: Do these correspondences suggest that the overt feminine accusative **-ta** suffix in Kambaata might have had an older masculine counterpart **-k** (or **-h**), reflexes of which surface in Hadiyya **-(k)icco**? That would allow the reconstruction of earlier ***-t** and ***-k** for both, with each surviving in one language.[138]

3.3.4.2 Triggers and Outcomes

The insertion of intervocalic **-ʔ-**, primarily observed in Hadiyya and Gedeo, functions as a key strategy for resolving vowel hiatus arising from (nominal) morphological processes (chiefly suffixation) or potentially from the historical elision of consonants. The specific environments and subsequent effects are synthesized in the following paragraphs.

Epenthesis **-ʔ-** is consistently triggered when two or more vowels become adjacent, i.e., across a morpheme boundary when a vowel-initial suffix is added to a vowel-final stem. First, simple **V+V** is found in numerous examples showing insertion between two vowels, such as Hadiyya *anno-ʔo* ← *anna* + **V** ‘father’ (pl), Gedeo *daana-ʔe* ← *daanalco* + **V** ‘elephant’ (pl), Hadiyya *aro-ʔo* ‘husband’, Gedeo *amaa-ʔa* ‘mother’ (acc). Second, complex vowel sequences

[138] See even the Sidaama parallels with as **-k** in Hadiyya *fuut’eʔe*, *fuut’ek-icco*, Kambaata *fuut’eeccu*, and Sidaama *futakko*, *futaho*, *futahe* (pl) ‘mole (animal)’. The double **uu** in Hadiyya and Kambaata might be compensatory upon fusion of the **t** with a glottal stop of unknown origin(§4.1.3) and degemination **tʔ** → **t’t’** → **t’**.

VV or **V́** also trigger epenthesis, but the source of the relevant vowel sequence is not always recoverable (original double vowels vs. suffixation). Hayward (1997) suggested that **-ʔ-** splits original long terminal vowels (**VV** → **VʔV**), and Tadesse Sibamo (2015) proposed vowel duplication (**V** → **VʔV**) before case marking in Hadiyya proper nouns. However, the variation in vowel quality, quantity, and accent surrounding epenthetic **-ʔ-** suggests that simple duplication or splitting is insufficient. It is often triggered by suffixation onto a stem ending in **V**. Examples potentially involving underlying **VV** or **V́** include Hadiyya *fellaʔa* (pl) 'goat', Gedeo *ollaʔa* 'neighbour' (cf. Burji *ollaa*), and Hadiyya *mooyyeʔe* 'mortar' (cf. Burji *mooyyee*).

In terms of the interaction with suffixation and stem vowels, the epenthesis process evidently involves adding vowel-initial suffixes (case, number, possibly derivational) to stems ending in a terminal vowel (TV) without necessarily deleting the TV. Assimilation often occurs between the TV and the suffix vowel, particularly progressive assimilation in Gedeo, where **V_{TV}** + **VV** suffix results in **V_{TV}-VʔV** (with the first **V** of the suffix assimilated to **V_{TV}**). The accent and syllable structure also play a role because epenthesis may be sensitive to accent placement, possibly serving to separate the accented suffix vowel from the unaccented terminal vowels (V_{TV}-V́), which later shifts the accent to the first vowel of the heavy syllable (**V́-ʔV**), i.e., the former TV. In Gedeo accusative pronominals, the **V́-VʔV** structure can be reduced to **-V́** by dropping the final unstressed **-VʔV** portion, with high pitch realized on the new final vowel. This suggests a connection between epenthesis, accent, and syllable weight/stress (§2.4).

Epenthesis also seems to resolve the hiatus created by consonant elision. Epenthetic **-ʔ-** frequently appears in Hadiyya/Gedeo where other languages show **k**, **w**, or **ø**. In Hadiyya, the underlying stem-final consonant (often **k**) surfaces before certain suffixes (e.g., singulative **-icco**), while **-ʔ-** appears before others. Example: Had *fellaʔa* (pl) vs. *fellak-icco* (sg). This suggests two possible analyses: (a) an underlying consonant (**k**, **w**) is elided intervocalically (**V_V**), and **-ʔ-** is subsequently inserted into the hiatus, or (b) the consonant and the epenthetic **-ʔ-** are independent; **-ʔ-** insertion is triggered by the vowel-initial suffix regardless of whether a consonant has been elided. The surfacing consonant comes from the stem or specific suffixes (like a velar suffix in Hadiyya). Hadiyya verbal morphology also shows **-ʔ-** occupying suffix slots where other consonants (like **-tt-**) are expected.

Therefore, **-ʔ-** epenthesis is primarily a hiatus-resolution strategy in Hadiyya and Gedeo (and in restricted cases in Kambaata/Sidaama), triggered by vowel

adjacency, especially across morpheme boundaries during suffixation. It interacts in complex ways with vowel quantity, accent, and assimilation, and potentially occupies slots created by historical consonant elision. The specific vowel sequences triggering it involve combinations of **VV** and **V́**, leading to different insertion patterns (**V?V**, **V?VV**, **VV?V**).

While **-ʔ-** epenthesis represents a prominent strategy for resolving vowel hiatus, particularly in Hadiyya and Gedeo, it is crucial to acknowledge alternative outcomes observed elsewhere in HEC and even within the same languages in different contexts. When sequences of vowels arise, typically through suffixation onto vowel-final stems ($\mathbf{V_{TV}}$ + **V**), languages like Burji, Sidaama, and Kambaata often exhibit outcomes involving vowel coalescence, resulting in accented short vowels (**V́**) or accented long vowels (**V́V**).

This coalescence pathway might occur directly ($\mathbf{V_{TV}}$ + **V** → **V́(V)**) or indirectly. An indirect route could involve an intermediate stage of epenthesis, not necessarily limited to **-ʔ-**. Given that glides and glottals are susceptible to lenition and elision in intervocalic positions (§3.3.3, rule ii), an earlier epenthetic glide (**-y-**) or glottal (**-ʔ-**) inserted to break hiatus could subsequently be lost. The loss of this intervening consonant could then trigger compensatory coalescence or accentuation on the remaining vowel(s). Evidence for such intermediate glide epenthesis resolving hiatus is indeed found in Burji, Sidaama, and Kambaata (§3.4.1.2).

Therefore, when analyzing instances of vowel hiatus across HEC, particularly those resulting from suffixation, both the conditioning context and the repair strategy may be language-specific. Three potential outcomes must be considered: (a) epenthesis of **-ʔ-** (predominant in Hadiyya, Gedeo), (b) epenthesis of **-y-** (common in Burji, Sidaama, Kambaata), and (c) vowel coalescence/accentuation (**V́** or **V́V**), potentially arising either directly from **V**+**V** or indirectly following the elision of an earlier epenthetic **-ʔ-**, or **-y-**.

In summary, vowel hiatus created by suffixation on vowel-final stems is resolved by means of different repair strategies across HEC. Hadiyya and Gedeo most consistently insert **-ʔ-**, while Burji, Sidaama, and Kambaata more often show glide epenthesis (**-y-**, **-w-**) or vowel coalescence with accent re-alignment. For reconstruction, the comparative pattern supports vowel-initial suffixation and stem-final (including terminal) vowels as the common conditioning context, but it does not require reconstructing **-ʔ-** as a proto-segment in longer stems. The distribution of **-ʔ-** epenthesis is therefore a

candidate for a shared innovation (subject to further testing against contact) and should be treated separately from inherited **ʔ** in monosyllabic stems.

Finally, the glottal stop participates in HEC morpheme boundary alternations in several distinct ways that should be kept distinct in reconstruction. As a derivational suffix, **-ʔ** is securely reconstructable for PHEC and remains productive across the family. In NHEC, the interaction of suffixal **ʔ** with stem-final consonants yields regular fusion patterns that motivate reconstructing a PNHEC morphophonological process of glottalization, even though Burji lacks the productive fusion outcomes. By contrast, intervocalic **-ʔ-** in longer stems is best treated as a later hiatus-repair strategy, concentrated in Hadiyya and Gedeo and competing with glide insertion and vowel coalescence elsewhere.

3.4 Sonorants

Sections 3.1-3.3 focused on processes affecting obstruents, with sonorants acting mostly as passive participants there (as targets or neighbors of obstruent alternations rather than the primary triggers). Their most active role in Chapter 3 is as epenthetic sounds. This section, therefore, focuses on sonorant insertion: epenthetic **-y-** (often realized as **-i-**) (§3.4.1), epenthetic **-n-** (§3.4.2), and epenthetic **-l-** (§3.4.3). The inserted segments adapt to their environments, with alternative realizations as **-w-**, **-m-**, and **-r-**, respectively.

3.4.1 Epenthetic -i-, -y-

The HEC languages insert epenthetic segments to resolve phonotactically illicit sequences. The most common strategy between consonants is **-i-** insertion, which breaks up impermissible clusters by creating an additional syllable. In intervocalic contexts, hiatus is often repaired by glide insertion (typically **-y-**), and surface **-i-** in some environments plausibly reflects this same repair strategy. These patterns are productive across HEC. In Hadiyya and Gedeo, intervocalic hiatus resolution is instead frequently realized as a glottal stop (§3.3.4).

3.4.1.1 Epenthetic -i-

Epenthetic **-i-** is the most common insertion strategy and is found throughout HEC. It occurs between (a) retained stem-final consonant(s) and a consonant-initial suffix (**(C)C**$_{i}$-i-**C**). Its primary function is phonotactic: It repairs impermissible consonant sequences, including clusters that are illicit because

of their segmental makeup (quality) or because they yield an impermissible consonant quantity (e.g., a three-consonant sequence).[139]

C_i + C → C_i-i-C

Below, segmentation shows only the stem-final consonant and the suffix to highlight the repair environment; the full lexical stem appears in the example.

Had	**ʃ- + -títte → ʃ-i-títte**	*ʃ-* 'kill' → *ʃ-i-títte* 'do not kill' (sg)
	k' + -cc → k'-i-cc	*mík'-* 'bone', *mik'-í-cc-í-cco* 'bone' (dimunitive-sgv)
Kam	**ʃ + -teen-óta → ʃ-i-teen**	*ʃ-* 'kill' → *ʃ-i-teen-óta* 'in order to kill'
	t + -n-óommi → t-i-n-	*fint-* 'list' → *fint-i-nóommi* 'we listed'
Sid	**ʃ + -tú → ʃ-i-tú**	*ʃ-* 'kill' → *ʃ-i-tú* 'she/they kill'
	b + -tú → b-i-tú	*sirb-* 'sing' + *-tú sirb-i-tú* 'she/they sang'
Ged	**ʃ + -siis → ʃ-i-siis**	*goʃ-* 'uproot' → *goʃ-i-siis* 'uproot' (caus)
Burji	**l + -ngạ → l-i-ngạ**	*k'al-coó* 'goat' → *k'al-í-ngạ* 'of goat' (gen)[140]

Sidaama inserts **-i-** before the reflexive suffix as in *af-* 'to find, get to know', *af-i-d'-* 'to have, get'; *has-* 'to look for' *has-i-d'-* 'to want'; *arg-i-d'-* 'to borrow (something that can be returned as it is)'; and *fooliʃʃ-i-d'-* 'to take a rest' (Kawachi, 2007, pp. 335–339, 324, 58). However, besides epenthesis, such **i** is a reflex of **y* 'say, do, be', e.g., in Gedeo *mar-* 'go' → *imar-* 'take' (lit. 'cause to go') and *dag-* 'come' → *idag-* 'bring' (lit. 'cause to come') (Eyob Kelemework, 2015, pp. 183–184).

3.4.1.2 Epenthetic -y-

Epenthetic **-y-** is less common than **-i-** and is found in Burji, Kambaata, and Sidaama.

V_i + V → V_i-y-V

[139] Data sources and further discussion: Hadiyya (Tadesse Sibamo, 2015, pp. 43, 263), Kambaata (Treis, 2008, pp. 61–62), Gedeo (Eyob Kelemework, 2015, pp. 76–77), Sidaama (Anbessa Teferra, 2000, p. 17), and Burji (Tesfaye Baye, 2015, p. 75).

[140] Compare the above with Gedeo *ʃiy-* 'kill' → *ʃiye ~ ʃie* 'he killed' (Eyob Kelemework, 2015, p. 36) and Burji *sey-* 'kill' → *sey-i-nnoó* 'killed' (perf) (kill.3SG-EP-PRV-TV) (Tesfaye Baye, 2015, p. 88). These Burji and Gedeo stems may be cases of double epenthesis (**-i-** and **-y-** (§3.4.1.2)) or include a support verb **y* 'to say, do, be' and epenthetic/phonetic **i** because the root is just the consonant **ʃ**. However, the vowel **e** of the Burji cognate remains puzzling.

As described by Tesfaye Baye (2015, p. 79) for Burji and Treis (2008, pp. 32–33) for Kambaata, it surfaces as **-y-** adjacent to the front vowel **i**, while adjacent to the back vowel **u** it surfaces as **-w-**. This alternation is best treated as assimilation, and comparable boundary-conditioned assimilation is widespread in HEC.[141]

	-y-	**-w-**
Kam	*k'ohi* → **k'o* + *i* → *k'oyi* 'marrow'	*k'ahu* → **k'a-* + *-u* → *k'a-w-u* 'small'
Sid	*dag-í* 'come' → **da* + *í* → *da-y-i* 'he came'	*dag-ummo* → **da-* + *-ummo* → *da-w-ummo* 'I came'
Burji	*saá* 'cow' → *se-y-igụ* (def.m), *sé-y-ị* (nom.m)	*d'ug-* 'drink' → **d'u-* + *-ann-oó* → *d'u-w-ann-oó* 'drink (perf)'

Most, if not all, instances of epenthetic **-y-** occupy the position of a dropped consonant: **h** in Kambaata,[142] **g** (← ***k**) in Burji and Sidaama, and ***ʔ** in Burji. It comes as no surprise that those consonants (mainly ***k**, ***h**, ***ʔ**, ***y**, ***w**) are dropped through the historical HEC rules and processes (see §§3.2, 3.3.3, and rule ii). Therefore, epenthetic **-y-** occurs after a stem-final vowel, i.e., $\mathbf{V_i_V}$, occupying the slot of a dropped consonant when attached to a vowel-initial suffix.[143]

Concerning the **g** and **w** in Sidaama and Burji, Anbessa Teferra's (2014, p. 78) analysis of **g** loss and glide epenthesis in Sidaama aligns better with the epenthesis patterns than Sasse's (1982, p. 18) and Tesfaye Baye's (2015, p. 79) proposal of free variation between **g** and **w** in Burji. The velar **g** is a voiced variant of the voiceless **k**. Historically, ***k** has been lenited and lost in post-vocalic position in HEC (§3.2.2). Hence, similar to Sidaama, the loss of the velar is also posited as an intermediate stage for Burji before epenthesis *d'ug-* → *d'u-* (loss of **g**) → *d'u-w* (epenthesis of **-w-**) 'drink'. The same is proposed for Kambaata **h** and **w**/**y**.

---

[141] Data sources and discussion: Kambaata (Treis, 2008, p. 39), Burji Tesfaye Baye (2015, pp. 79, 181), Sidaama (Anbessa Teferra, 2014, p. 78), Gedeo (Eyob Kelemework, 2015, p. 36)

[142] See PNHEC [660]**k'oʔ-* 'marrow of bones' → Had *k'oʔo* 'marrow, bone marrow', Kam *k'o(o)ʔa* 'marrow, bone marrow' (Hudson, 1989, p. 96), Sid *k'oʔè* 'marrow, bone marrow', *k'ore* 'marrow of bones, cleverness'.

[143] Similar epenthesis is found in other Cushitic languages, e.g., Ehret (1987, p. 9) notes that "Agaw verbs with original stem final consonants which have been lost, normally pharyngeals and glottals, have often added *-w, occasionally *-y, in the place of the missing consonant, and these additions can be understood as the *w and *y extensions used as place holders rather than semantic modifiers".

Burji examples from Tesfaye Baye (2015, p. 79) show underlying ***V?*** or ***VC** → ***V-y-V** → **VV** → /__# surfacing as **V-y** when followed by a **-CV** suffix, confirming the existence of an underlying (lost) ***C**, e.g., **saʔ-* ‘cow’, **waʔ-* ‘water’ → **we-y-íg̣u̯* (def), *wé-y-i̯* (nom.m) (§3.3.3).[144]

Hadiyya and Gedeo use different insertions. Hadiyya has no epenthetic **-y-**, and instead uses **-ʔ-** epenthesis productively between vowels. Gedeo also has a preference for epenthetic **-ʔ-** intervocalically (§3.3.4).

As discussed in §4.2.4, the HEC languages have what is known in the literature as compound or phrasal verbs (Hudson, 1976, p. 273; Crass, 2007, p. 30; Treis, 2024). One of the two support verbs used in such formations is **y* ‘to say, do, be’, which is identical to the epenthetic **-y-** and also follows the stem (see §4.2.4.1). In such formations, the glide **y** (intervocalic) and **i** (interconsonantal) are not epenthetic but rather support verbs, which may apply to verbal forms, such as, e.g., Gedeo *ʃiy-* ‘kill’ → *ʃiye ~ ʃie* ‘he killed’. However, a vulnerable sound like the glide **y** is often elided following the verb stem (rule ii) (e.g., in Hadiyya (Hudson, 1976, p. 273)) or fused with the preceding verb stem, becoming difficult to distinguish from other verbal constructions, e.g.:[145]

Libido	**k’ + y → k’y/k’**	*bak’imma* ‘accroach’ → *bak’(y)imma* ‘awaken’
	l + y → ly/l	*k’ullimma* ‘be bald’ → *k’ul(y)imma* ‘be afraid’
Burji	**r + y → r**	*gagari yed’a* → *gagared’a* ‘I have a cold’
	V + y → Vy	*k’ufaa* ‘cough’ → *k’ufey-* ‘to cough’
	y + s → ys	→ *k’ufeys-* ‘to cause to cough’

The same applies to K’abeena (Crass, 2005, p. 228) and Alaaba (Schneider-Blum, 2007, p. 277). In Kambaata, likewise, the root consonant of “*y-* ‘say’ is often hardly audible in non-careful speech” (Treis, 2024, p. 162).

The remaining epenthetic **-n-** (§3.4.2) and **-l-** (§3.4.3) do not occur intervocalically or interconsonantally. Rather, they occur as the first members of consonant clusters, consistent with their historical position in complex syllable codas (§2.3.3).

[144] One can alternatively interpret the **y** as a manifestation of the proposed older nominative suffix ***-y** (§4.2.4.1) with the **-i-** being epenthetic.
[145] Data sources: Libido (Crass, 2007, p. 30), Sidaama (Kawachi, 2007, p. 206), Burji (Hudson, 1976, p. 264).

3.4.2 Epenthetic -n-

The epenthesis of **-n-** is an intriguing pattern in the historical phonology of HEC. It involves inserting a nasal **-n-** immediately before non-sonorant consonants. Functionally, this insertion appears to be phonotactic: It repairs marked consonant sequences by supplying the medial consonant position (**C_2** slot) in **C_2C_3** clusters, including both heterorganic clusters and clusters created by gemination. The attestations are scattered across the HEC languages, concentrated especially in Gedeo and Hadiyya, and attested in the lowest numbers in Sidaama.

[329]	(C)C	-n-	Gloss
PNHEC	***reħ-ʃ-**	**-**	‘corpse’
Had	**leeʃa, leeʃʃa**	**-**	‘corpse’
Kam	**re(e)ʃa**	**-**	‘corpse’
Sid	**reeʃʃa, reeʃa**	**-**	‘corpse’
Ged	**-**	**reenʃa**	‘corpse’

The lack of any other retained **n** in the rest of HEC supports the epenthesis analysis via gemination, as the correspondences show. Otherwise, since this is an example of a deverbalized stem, an alternative explanation would be to assume the **n** is part of the deverbative suffix as seen in §4.1.2.

[494]	(C)C	-n-	Gloss
PHEC	***k’ob-, *k’ok’~k’obb-**	**-**	‘to surround, to bend’
Burji	**k’ok’obaa**	**-**	‘to prepare temporary shelter’
PNHEC	***k’ob-, *k’ok’~k’obb-**	**-**	‘to surround, to bend’
Had	**k’ubb-**	**-**	‘to bend down’
Kam	**k’ubb aʔ-**	**-**	‘to bend, be(come) circular’
Sid	**k’ok’k’obi**	**-**	‘fence in’
Ged	**-**	**k’ok’k’onba**	‘be surrounded’

This stem is partially reduplicated, and the final **b** has been geminated before the degemination and **-n-** epenthesis in Gedeo.

[190]	(C)C	-n-	Gloss
PNHEC	***arr-**	**-**	‘sun’
Had	**-**	**ellinco**	‘sun’
Kam	**arriiccu-ta**	**-**	‘sun’
Sid	**arriʃʃo**	**-**	‘sun’
Ged	**ariʃʃo**	**-**	‘sun’

This is an instance of epenthetic **-n-** in Hadiyya corresponding to geminates in the rest of NHEC. The same applies to the following set [478]:

[478]	**(C)C**	**-n-**	**Gloss**
PHEC	***kott-**	**-**	‘hoof, foot’
Burji	**kottee**	**-**	‘hoof, track’
PNHEC	***kott-**	**-**	‘hoof, foot’
Had	**-**	**konteʔe**	‘hoof, foot’
Kam	**kotti-ta**	**-**	‘foot step, track’
Sid	**kotte**	**-**	‘hoof’
Ged	**kotte**	**-**	‘hoof’

The gemination might have been the result of singular **kot-* receiving a reduplicated **t** in the plurative formation **kot~t-* ‘pair of feet’, with the latter form surviving except for Hadiyya, where ***tt** is degeminated and compensated for by **-n-** epenthesis.

[514]	**(C)C**	**-n-**	**Gloss**
PHEC	***ebb-**	**-**	‘to bring’
Burji	**-**	**amb-**	‘bring’
PNHEC	***ebb-**	**-**	‘to bring’
Had	**eeb-**	**-**	‘bring, marry’
Kam	**eeb-**	**-**	‘bring, marry’
Sid	**ʔabb-**	**-**	‘bring’

The epenthetic **-n-** became **-m-** via assimilation to the following labial in Burji. The adaptation of **nC** clusters shows variation. In Gedeo **nb** becomes **nb** [494] or **mb** [495] below. This variability suggests the process is not fully regular and may reflect multiple historical layers of phonotactic development.

Potentially, additional examples of **-n-** epenthesis demonstrate the complexity of this phenomenon, where sometimes there is no clear corresponding geminate to motivate the **-n-** insertion:

Hadiyya	**[19] sadeento**	‘eight’	← ***sazzeett-**
	[45] fiink’ees-	‘to whistle’	← ***fiik’ ʔ**
	[723] seenk-	‘to beautify’	← ***sek-**
Kambaata	**[91] undulumma**	‘mortar’	← ***udulunm-a**
	[390] iinc’aʔ-	‘stretch’	← ***hiit’t’-**
Gedeo	**[495] lumo**	‘big’	← ***lumb-**
	[113] dammara	‘to add’	← ***dambar-**
Sidaama	**[29] ʔamadi**	‘hold, catch’	← ***amb-**
Burji	**[351] t’uunt’-, t’unt’-**	‘to absorb’	← ***t’uuCt’-**
PNHEC	**[46] *ink’-, *ink-**	‘tooth’	← ***ik’k’-, *ikk-**

The sequence of epenthetic **-n-** and the following **C** usually corresponds to geminates (**ʃʃ** [329], **tt** [478], **t’t’** [390], **kk**/**k’k’** [46]) and complex clusters (***Ct’** [351]). The gemination itself is primarily a result of a morpheme boundary alternation process. For instance, a number of those stems with **-n-** epenthesis are followed by a glottalized consonant from an intermediate glottalization and gemination process discussed above (§3.3.1). Non-glottalized geminates, on the other hand, may have developed from the full assimilation of a stem-final obstruent with various suffixes, including, for instance, the nominal derivational ***-ʃ** in [329] and [190], or similar fossilized suffixes (§§3.1.1, 3.3.1). Stem-final consonants are also geminated through other derivation processes, such as before the Plurative *-uw-a (§4.1.1.2) and in the comparative set [478] above.

Comparable **-n-** insertion has been reported for Gurage (Leslau, 1979, p. 1vi) and other Semitic languages. At present, however, it remains unclear whether contact has played any role in shaping the HEC pattern, and if so, at what stage(s) and in which subareas.

Some apparent cases of **-n-** epenthesis may instead reflect other historical developments, such as the metathesis of a suffix ***n** (Kawachi, 2007, p. 46; Eyob Kelemework, 2015, pp. 79–80). This possibility underscores the need to distinguish phonotactic repair from morphology-driven restructuring; broader comparison with other East Cushitic languages and neighboring non-Cushitic languages may help clarify the origin and trajectory of the pattern.

3.4.3 Epenthetic -l-

A further, less common epenthetic consonant, **-l-**, is proposed to account for puzzling cases of isolated **l**/**r** before stem-final consonants, primarily in Gedeo (and also reported for Gurage (Leslau, 1979)). The rarity of this pattern suggests that **-l-** epenthesis is relatively recent compared to other insertion

processes, and it appears to occur in the same structural environment as epenthetic **-n-** (§3.4.2).

[460]	**(C)C**	**-l-**	**Gloss**
PNHEC	***beet-**	**-**	‘son, child’
Had	**beéto̩**	**-**	‘boy, son, child’
Kam	**beetu**	**-**	‘child (son/daughter)’
Sid	**beetto**	**-**	‘son, boy, child, daughter, girl’
Ged	**-**	**belto, berto**	‘boy’

The corresponding geminate is attested only in the Sidaama cognate and is compensated for by epenthesis in Gedeo. Both **-l-** and **-r-** realizations of epenthetic **-l-** occur, without a clear conditioning environment.

Set [460] shares the same root with PNHEC [741] **beet-, *beet-beet-* ‘grandchild’, which also continues with geminate in Sidaama as *beettu-beetto* and epenthetic **-l-** in Gedeo *beltim-belto* ‘grandchild’. To explain the Gedeo reflexes, additional insertions need to be proposed: after **-l-** epenthesis **belt-belt-*, **-i-** epenthesis **belt-i-belt-*, and **-n-** epenthesis **belt-i-n-belt-* occurred, potentially in a different order.

[498]	**(C)C**	**-l-**	**Gloss**
Amharic	**wəʃʃa**	**-**	‘dog’
Burji	**wacco, waʃʃa, woccoo, woʃa, woʃʃi**	**-**	‘dog’
Had	**waʃicco, wiʃicco, wuʃa**	**-**	‘dog’
Kam	**woʃa-ta, woʃ-iccu**	**-**	‘dog’
Sid	**waʃ-icco, waʃʃa, woʃ-icco, woʃa**	**-**	‘dog’
Ged	**-**	**wolco, warʃa**	‘dog’

This example might be a loan from Amharic, hence the variation in the root vowel.

[299]	**(C)C**	**-l-**	**Gloss**
PNHEC	***zaane-i-ʃʃ-**	**-**	‘elephant’
Had	**daaneecc**	**-**	‘elephant’
Kam	**zaneeccu**	**-**	‘elephant’
Sid	**daaniicco**	**-**	‘elephant’
Ged	**-**	**daanalco**	‘elephant’

Similar to [190] *ellinco* ‘sun’ in Hadiyya, with epenthetic **-n-**, the geminated singulative in the rest of NHEC corresponds to epenthetic **-l-** in Gedeo in [299]. This might reflect a placeholder of the old stop that triggered the original

affrication posited for ***-ʃ** → **-c** /**C**$_{\textbf{[+stop]}}$**_** in §3.1.1.5. Another example is Gedeo [708] *moo?ralco*, *mood'aco* 'thief', with reflexes of the other verbal derivation suffix (deverbative) ***-ʃ** as seen in Kambaata *moga?-aancu*, Sidaama *moor-aanco* from PHEC [362] **moor ʔ* 'to steal'.[146]

[111]	**(C)C**	**-l-**	**Gloss**
PHEC	***baʔ-is-**	**-**	'to spoil, extinguish'
Burji	**baas-**	**-**	'to take out, bring out; extinguish fire'
PNHEC	***baʔ-is-**	**-**	'to spoil, extinguish'
Had	**baʔ-is-**	**-**	'extinguish'
Kam	**baʔ-is-**	**-**	'spoil, waste; extinguish'
Sid	**baʔi**	**-**	'perish, extinguish, be misled'
Ged	**-**	**ball-ess-**	'extinguish'

The gemination of the epenthetic **-l-** in this example might be due to assimilation to the stem-final or attached suffix consonant.

Burji has two examples where **-r-** is suspected to be an epenthetic.

> PHEC [46] **ik'k'-* (← **ilk-ʔ-*), **ikk-* 'tooth' → Burji irk'a, PNHEC **ink'-*, *ink- → Had *ink'e*, Kam *ink'u-ta*, Sid *hinko*

Burji **-r-** in [46] *irk'a* 'tooth' is analyzed as epenthetic rather than a change of **l** to **r** because the latter is not a regular change in Burji. This same stem shows epenthetic **-n-** in PNHEC **ink'-, *ink-* 'tooth'. The loss of the earlier ***l**, Burji **-l-**, and PNHEC **-n-** insertion in this stem are triggered by morpheme boundary alternation processes to resolve the **CCC** cluster upon the stem-final ***k** fusion and gemination into **k'k'** (§3.3.1.3) and **kk** (§3.1.1).

> PHEC [380] **hik'k'-*, **hek'k'-* 'have hiccups' → Burji *hirc'immeed'-*, PNHEC **hik'k'-*, **hek'k'-* → Had *hik'k'aʔ-*, Kam *hik'k' aʔ-*, Sid *hek'k'i yaa*, *hek'k'i-ass-is-*, Ged *hek'k'eʔmef-at-* (← Oromo), *hek'k'eʔm-is-* (see Oro *hirk'inf-* 'have hiccups')

The other instance of **-r-** in Burji [380] *hirc'immeed'-* 'have hiccups' might be due to Oromo *hirk'inf-* 'have hiccups', in which the same epenthesis hypothesis might apply when compared with Konso *hik'k'ota* 'hiccup'. The

[146] The **n** in the Sidaama and Kambaata cognates might be a concatenated suffix **-am** as discussed in §4.1.2. The PHEC reconstruction of [708] is **moor-k-* 'thief'. The **k** is probably an agentive or deverbative suffix equivalent to **t** in Hadiyya, but requires further inspection. See Burji *moor-g* (m) 'who steals' (Hudson, 1989, p. 152).

palatalized **c'** in Burji might be due to an independent morpheme boundary alternation process (see §§3.1.2 and 3.1.2.2).

With **-l-** epenthesis included, the full class of sonorants is attested as epenthetic material in HEC. However, Eyob Kelemework's (2015, p. 36) analysis of Gedeo **y**, **l**/**r**, and **n** treats them as assimilatory realizations of an underlying **y** before alveolar stops and alveolar nasals, respectively; i.e.:

y → { **l** or **r** /__**C**[+alveolar stop] ; **n** /__ **C**[+alveolar nasal] ; **y** /__**V** }

That would rather mean the three are epenthetic, with **l**/**r** being the most recent, since it is not common outside of Gedeo. One could arguably assume the same for the free variation between **d'** and **r** intervocalically, e.g., *fad'acco* and *fararcco* 'horse', and **l** word-initially, e.g., *d'uga* and *luga* 'truth', *d'ufo* and *lufo* 'fart' (Eyob Kelemework, 2015, pp. 37–38). The word *d'uga*, *luga* is not attested in any comparanda in the current study, and *d'ufo*, *lufo* (with ***s** → **f**) is from Oromo, while **d'** in *fad'acco*, *faracco* is from an original ***r**. Such alternation has not been attested for inherited ***l** or ***r**. Therefore, it is possible to argue that the **-l-** is recent, and that it has become part of the **d'**, **ʔ**, and **ø** alternation group. Applied analogically to the **d'** in initial position and the original ***r**, e.g., *fad'acco* 'horse'.

Eyob Kelemework (2015, pp. 89, 192) also claims that the implosive **d'** of the middle suffix is optionally replaced by **ʔ** when the stem-final is a sonorant or glottal stop and **l** if the stem-final is an obstruent, with subsequent metathesis; e.g, the stem *af-* 'get, have' becomes *af-ed'-* with the middle *-ed'*, *alf-* with the middle **-l**, and both could be found on the same root in *alf-ed'*. The same goes for *goʃ-* 'uproot', which becomes *goʃ-ed'-*, *golʃ-*, and *golʃ-ed'-*. That means the environments specified earlier, **d'** → **r** intervocalically and **l** initially, do not account for the claimed suffix **d'** change to **l**. Therefore, it is more efficient to assume that it is the common Gedeo **-l-** epenthesis. Another incentive for the alternation could be due to the origin of **d'** as a borrowed sound, which might pose pronunciation difficulty for speakers of some dialects. Hudson (1989, p. 7) links this dialectal Gedeo change to the change of Sidaama **d'** to **ʔr** in intervocalic position, e.g., *had'-* 'go' → *haʔrí* (perf) 'he went' (Anbessa Teferra, 2000, p. 26) and *afid'-* 'to have' *afiʔra* (infinitive) 'he has' (Kawachi, 2007, p. 47). However, Sidaama **ʔr** is different from Gedeo **r**.

In summary, consonant insertion (sonorants and glottals) in these contexts serves syllable-structure requirements: It supplies a minimal consonantal margin intervocalically and can also occupy the position left by (i) a weakened or lost consonant cluster (**cC**) or (ii) the first member of a degeminated (**CC**) sequence.

The various manifestations of the inserted sounds are summarized below (epenthetic **-ʔ-** is treated in detail in §3.3.4).

Epenthesis	Realizations	Input	Outcome
-y-	**-i-**	*C_i+C	→ C-i-C
	-y-	*$V_{[-back]}C_i$+V or *VC_i+$V_{[+back]}$	→ V-y-V
	-w-	*uC_i+V or *VC_i+u	→ *u-w-V or V-w-u
-n-	**-n-**	*$CC_{i[-labial]}$	→ -n-C
	-m-	*$CC_{i[+labial]}$	→ -m-C
-ʔ-	**-ʔ-**	*$V_{i/TV}$+V or *$VC_{i/TV}$+V	→ $V_{i/TV}$-ʔ-V
-l-	**-l-, -r-**	*CC_i	→ lC_i, rC_i (Gedeo)

The schematic summary list above abstracts away from language-specific surface details in order to highlight recurring repair strategies and their conditioning environments. Two structural domains are especially prominent. First, boundary-driven repairs in stem+suffix combinations create an additional syllable to break up impermissible consonant sequences, most commonly through vowel/glide material (e.g., interconsonantal **-i-**, here treated as compatible with analysis as an underlying glide). Second, hiatus repair in vowel sequences is resolved either by glide insertion (**-y-**, with **-w-** as a conditioned realization next to back vowel; both **-y-** and **-w-** are attested between (or adjacent to) **a**) or, in a more restricted subset of languages, by glottal stop insertion.

The distribution of these strategies is uneven across HEC and is therefore informative for the comparative discussion. Interconsonantal epenthesis is broadly productive and appears in multiple morphological contexts, but individual instances must be distinguished from support-verb morphology (**y* ‘to say, do, be’ → **i**) where the surface segment **i** is not straightforwardly phonotactic. Intervocalic **-y-** (and **-w-** as its conditioned realization) is concentrated in Sidaama, Burji, and Kambaata, while intervocalic **-ʔ-** is concentrated in Hadiyya and Gedeo. The consonantal insertions **-n-** (with **-m-** as a labial-conditioned realization) and **-l-**/**-r-** occur as cluster-initial material and are comparatively rarer, suggesting more limited diffusion or later innovation.

The input patterns preceded by an asterisk in the above list should not be read as direct proto-forms. They may represent intermediate stages that were subsequently remodeled by assimilation and other boundary processes. Based on the current comparative evidence presented in §3.4, only ***-y-** (together with its conditioned realizations) is proposed as a secure PHEC epenthetic segment. The remaining insertions, such as **-ʔ-** in hiatus repair and **-l-** in cluster contexts, are better treated as later, language-specific (or contact-induced) developments unless future comparison establishes regular correspondences and consistent cognate-set behavior.

Interconsonantal **-i-** epenthesis is uncontroversial in most cases, except where there is uncertainty about whether the segment reflects the support verb ***y** 'to say, do, be' (→ **i**). Intervocalic **-y-** and **-ʔ-** overlap in distribution across the HEC languages (**-y-** in Sidaama, Burji, and Kambaata; **-ʔ-** in Hadiyya and Gedeo). Analyses that avoid interpreting the non-intervocalic epenthetic **-C-** (**-n-**, **-l-**, **-ʔ-**) as insertion are difficult for two reasons. First, the relevant correspondences involve singletons (**C**), geminates (**CC**), and clusters (**cC**). Second, the same cognate set typically does not show consistent epenthesis across languages (with only one or two exceptions). This distribution points to a recent, language-specific development, and it prevents establishing the regular correspondences required to reconstruct all such epenthetic consonants to the proto-level. Thus, based on the available evidence, only ***-y-** and its different realizations are proposed for PHEC, and ***-n-** is provisionally posited for PNHEC.

The following section (3.5) synthesizes these obstruent alternations and sonorant repairs into a comparative summary and a set of reconstruction-oriented conclusions.

3.5 Summary and Conclusions

This section synthesizes the morpheme boundary alternation processes documented in the preceding sections. It summarizes their distribution across the HEC languages and their reconstructability to PHEC or PNHEC. The summary table presents each major process with its attestation pattern; the numbered points below highlight key findings. Some interpretations remain tentative pending further investigation of the comparative data.

Notation in the table: * next to a numbered point indicates the process is reconstructable to PHEC; a subscript on a consonant or on C_i indicates stem/root-final position; + indicates concatenation of segments; / indicates

language-specific variants conditioned by environment; and - means the specific process is not attested for the language.

	Process	Input	Burji	**PNHEC**	Had	Kam	Sid	Ged
1.	Assimilation (pron) **-t**	C_i[-son]+t	ʃʃ/cc/nd	***C_iC_i**	C_iC_i	C_iC_i	C_iC_i	C_iC_i/nd/ld
2.*	Assimilation (pron, plv) **-n**	C_i[+son]+n, s+n	C_iC_i, s+n → s-i-n	***C_iC_i, *ns**	C_iC_i, s+n → ss	s+n → ss	C_iC_i, s+n → ns	C_iC_i, s+n → ns
3.	Assimilation **(verbal) -s**	C+s	ss	***ss**	-	ss	ss/ʃʃ	-
4.*	Affrication (nominal) **-ʃ**	*ʃ	ʃ/s/c/cc	***ʃ**	ʃ/ʃʃ/c/cc	ʃ/ʃʃ/c/cc	ʃ/ʃʃ/c/cc	ʃ/c/cc/j
5.	Palat. (verbal) **-s**	C_i+*-s	-s	**-**	-	-ʃ	-ʃ	-ʃ/-c
6.	Palat. (pron) **t_j**	Ct_i+i, Vt_i+i	c/ʃ	**-**	-	-	-	-
7.	Palat. (pron) **k_j**	k_i+i	ʃ	**-**	-	-	-	-
8.	Elision **$ʔ_i$**	$Vʔ_i$	VV/y	***ʔ**	ʔ	ʔ	ʔ/ø	ʔ
9.*	Lenition **b_i**	Vb_i+V	β/w	***w**	w	w	β/w	β/w
10.*	Lenition **k_i**	*Vk_i+V	h/(ø)	***k/*h**	h/ø	h/ø	k/h/ø	k/h/ø
11.	Fusion	$C_{i[+son]}$+ʔ	-	***C'C'**	C', C'C'	C', C'C'	C', C'C'	C', C'C'
12.	Metathesis	$C_{i[+son]}$+ʔ	-	***ʔ$C_{i[+son]}$**	ʔ$C_{i[+son]}$	ʔ$C_{i[+son]}$	ʔ$C_{i[+son]}$	ʔ$C_{i[+son]}$
13.*	Epenthesis **-ʔ-, -i-, *-y-, -n-, -l-**	V_V, C_C, CC	-i-, -y-, -n-C	***-y-, *-n-**	-i-, -ʔ-, -n-C	-i-, -y-, (-ʔ-), -n-C	-i-, -y-, -n-C	-i-, -ʔ-, -n-C

Table 6: *Summary of Morphophonological Processes*

The morphophonology summarized above is supported by a small set of recurring boundary processes affecting stem-suffix combinations. Assimilation is especially clear in pronominal and derivational morphology. The 2nd/3rd-person suffix ***-t** shows regular assimilation to a preceding obstruent, with partial voicing assimilation in Gedeo (e.g., **nt** → **nd**, **lt** → **ld**). In Burji, gemination can be accompanied by palatalization (e.g., **tt** → **cc**/**ʃʃ**) when followed by the high front vowel **i**. The plurative and pronominal suffix ***n** likewise assimilates fully to a preceding sonorant (except in Kambaata). Assimilation of ***n** to a stem-final ***s** yields ***sn** → **ss**/**ns**, and Burji resolves clusters by inserting an epenthetic vowel after the stem-final consonant. Stem-final obstruents may also assimilate to the causative suffix in language-specific ways, yielding **s-s** where the causative surfaces as **-s** (Burji, Kambaata, Sidaama) and **-ʃ-ʃ** where the causative surfaces as **-ʃ** (Sidaama). Finally, the nominal derivational suffix is best reconstructed as PHEC ***-ʃ**, with an affricated allomorph ***-c** conditioned (at least historically) after stops and later generalized; Burji reflects ***-ʃ** as **ʃ** and **s**, while Gedeo additionally shows a voiced affricate **j** in restricted environments.

Palatalization interacts with several of these boundaries. Causative ***-s** palatalizes to **-ʃ** in Kambaata, Sidaama, and Gedeo. In Burji, stem-final ***t** palatalizes to **c** before **i**, and to **ʃ** intervocalically before **i**. Burji also shows palatalization of stem-final ***k** to **ʃ** before **i**.

Lenition targets intervocalic stops. Intervocalic ***b** lenites to **w** (and to **β** in Burji, Sidaama, and Gedeo), and this **w** can subsequently be lost. Intervocalic ***k** lenites to **h**, which can likewise be lost.

Other processes include elision, fusion, metathesis, and epenthesis. The loss of non-initial **ʔ** is widespread across HEC but more prevalent in Burji, where it is often accompanied by compensatory effects (especially vowel doubling and sometimes glide addition), so the absence of surface **ʔ** should be treated as weak negative evidence in comparative reconstruction. By contrast, two boundary processes involving suffixal **ʔ** are more restricted and therefore more informative for subgrouping: in NHEC, a stem-final obstruent can fuse with a following **ʔ** to produce a glottalized geminated consonant, and sonorant + **ʔ** metathesis (**C**$_{[+son]}$**ʔ** → **ʔC**$_{[+son]}$) is likewise restricted to NHEC. In Sidaama and Gedeo, the specific sequence **ʔr** may further alternate with a glottalized coronal **d’**(**d’**), plausibly reflecting a contact-induced reinterpretation of metathesized **ʔr** utilizing the fusion process. Finally, impermissible vowel and consonant sequences are commonly repaired by inserting **-y-** or **-n-** respectively, while Hadiyya and Gedeo sometimes use **-ʔ-** instead of **-y-**.

The major morpheme boundary alternation processes can be viewed as a chain of interrelated strategies of feature-changing (assimilation and fusion) that produce gemination and glottalization, segment loss (elision) that produces hiatus, segment insertion (epenthesis) that repairs vowel hiatus and fills consonant slots, and segment reordering (metathesis) that repairs impermissible sequences. For instance, the assimilation and fusion processes yield geminated (glottalized) consonants. These geminated consonants sometimes get degeminated and repaired by an epenthetic consonant, yielding consonant clusters, or compensated for by vowel doubling. The chain of actions and reactions across the different processes operates within the structural phonological constraints observed in the non-initial segments, as discussed in Chapter 2 (§2.3.3). This includes the preference for a simple (singleton) coda (which motivates occasional degemination), the fact that a complex coda (**cC**) must be a sequence of sonorant + obstruent (sequences of morphological obstruents call for full assimilation and resulting gemination (obstruent + obstruent) repaired by selecting a sonorant for epenthesis to settle on the preferred sonorant + obstruent sequence), and the ban on tripartites (**CCC** and **VVV**) and heterogeneous **VV** in addition to the recognition of underlying consonants (which trigger epenthesis or, rarely, reduction).[147]

This chapter has shown that many generalizations about HEC phonotactics and non-initial segmental structure become most transparent at morpheme boundaries, where otherwise latent contrasts and constraints are forced into interaction. In this sense, Chapter 3 is a direct continuation of Chapter 2: It expands on the structural constraints on syllable boundaries and segment sequences. This chapter demonstrates how those constraints are actively negotiated by morphology through recurrent alternations such as assimilation, elision, epenthesis, fusion, and metathesis. Because these processes systematically reshape surface forms, they are not merely “post-lexical repairs”; rather, they constitute a primary locus where evidence for reconstructing non-initial consonants and consonant clusters can be recovered and interpreted.

At the same time, the patterns surveyed here provide an essential bridge to the subsequent chapters. Chapter 4 turns to morphological structure more directly, including the affixes and paradigmatic environments that trigger the

[147] Fusion/glottalization occurs in both roots ending in **C** and **CC**, and their patterns exhibit an interplay of various phonotactics strategies, e.g., the few stems ending in **CC** simply become glottalized **C’C’** when **ʔ** is attached to avoid an impermissible sequence of three segments. The same applies to instances of the derivational **y**, but without glottalization or even gemination.

alternations at morpheme boundaries documented in this chapter, while Chapter 5 evaluates the implications of shared innovations and distributions across the HEC languages on subgrouping . Taken together, Chapters 2–4 establish the phonological and morphological background needed to distinguish broadly inherited tendencies from branch-specific developments, allowing the historical narrative and subgrouping argument to be stated on firmer, cumulative evidence.

4 Morphology

This chapter 4 reconstructs the nominal and verbal morphology of Highland East Cushitic and evaluates the distribution of morphological markers across the five HEC languages. The analysis follows the comparative framework established in Chapter 1 (§1.5) and is integrated with the phonological correspondences and reconstructions developed in Chapter 2. Each proposed marker is assessed in terms of its attestation, its allomorphy, and the environments by which it is conditioned, complemented by Chapter 3, along with its reconstructability to PHEC or to the PNHEC node.

Morphology is central to the interpretation of the boundary alternations documented in Chapter 3 because many processes are triggered by specific suffixes and stem classes. By making the relevant morphemes explicit and reconstructing them, this chapter provides a tighter account of why particular sandhi patterns recur across languages and why others are restricted to individual branches in HEC. The chapter also contributes evidence to the internal classification developed in Chapter 5 by identifying shared morphological innovations and by distinguishing inherited markers from branch-specific developments.

This chapter is organized as follows: Section 4.1 treats noun morphology, including number marking, nominal extensions, genitive constructions, and gender and case markers, with a separate subsection on personal pronouns. Section 4.2 turns to verb morphology and reconstructs the major derivational markers and support verbs. Section 4.3 synthesizes the reconstructed morphemes in a comparative summary and highlights the markers that are reconstructable.

4.1 Noun Morphology

This section surveys the main categories of noun morphology in HEC and reconstructs the markers that can be projected to PHEC or PNHEC. The discussion begins with number marking (§4.1.1) and nominal extensions (§§4.1.2, 4.1.3), then turns to genitive constructions (§4.1.4) and to gender and case morphology (§4.1.5). The final subsection addresses personal pronouns (§4.1.6).

Throughout, the aim is to separate broadly inherited patterns from branch-specific developments. For each marker, the analysis identifies its formal exponent(s), the environments that condition its allomorphy, and its

distribution across the five languages. Where nominal morphology triggers alternations at morpheme boundaries, cross-references to Chapter 3 are provided to contextualize the morphological description within the detailed discussion of the boundary processes.

4.1.1 Number

Highland East Cushitic number morphology distinguishes formally derived categories from functional ones. I use the terms singulative and plurative (Dimmendaal, 1983, p. 223; Corbett, 2000, p. 17) to refer to derived forms carrying overt number suffixes, whether marked against a base or against each other. Nouns may thus be formally unmarked, singulative, or plurative (see §1.5.2; Eyob Kelemework, 2015, p. 103), while singular and plural are reserved for function reference (Treis, 2008, p. 131).

4.1.1.1 Singulative

Alongside the suffix ***-ʃ*** (realized as ***-ʃo*** with a terminal vowel), Hudson (1976, p. 251; 2006, p. 296) identifies an HEC singulative ***-cco***, appearing as ***-co*** after sonorants and ***-cco*** after obstruents. Comparative evidence supports reconstructing these not as distinct suffixes, but as a single conditioned set. The proposed PHEC singulative allomorphs and their reflexes are summarized below.[148]

	PHEC	PHEC allomorphs	HEC language allomorphs
PHEC	***-ʃ →***	***-ʃ, *ʃ-ʃ, *-i-ʃʃ, *-c***	
Had	……		**-i-cco, -co**
Kam	……		**ʃʃu/ccu, -i-ccu**, **-cu**, **-ee-ccu**
Sid	……		**-ʃʃo**, **-i-cco**, **-co**
Ged	……		**-ʃo**, **-i-cco**, **-co**, **-jo**
Burji	……		**-ʃo, -ʃaa**, **-i-ccoo, -coo**

The distribution strongly supports a single singulative morpheme with phonologically conditioned allomorphs rather than multiple unrelated suffixes. The allomorphy is predictable: ***ʃ-ʃ*** occurs after fricatives (via fusion); ***-i-ʃʃ*** arises with epenthetic ***-i-*** after geminates or clusters; ***-ʃ*** marks stem-final sonorants; and ***-c*** follows stops. As argued in §3.1.1.5, the affricated allomorph ***-c***, originally restricted by context, appears to have expanded in the modern HEC languages at the expense of the post-alveolar allomorphs.

[148] Data sources and discussion: Burji (Tesfaye Baye, 2015, p. 95), Sidaama (Anbessa Teferra, 2000, p. 27), Hadiyya (Tadesse Sibamo, 2015, p. 48), Kambaata (Treis, 2008, p. 132), and Gedeo (Eyob Kelemework, 2015, p. 103).

In Gedeo, the additional singulative form **-jo** is treated here as a restricted voiced realization of the affricated allomorph (cf. §3.1.1.5 (Affricated **j**)), not as an independent singulative suffix.

A representative singulative example is PNHEC [190] **arr-* 'sun'. Sidaama *arriʃʃo* and Gedeo *ariʃʃo* 'sun' retain the post-alveolar allomorph ***-i-ʃʃ** in the environment of a geminate, whereas Hadiyya shows an affricated realization in *ellinco* 'sun'. Singulative marking is infrequently used in Burji (Hudson, 1976, p. 251).

4.1.1.2 Plurative

Plurative marking in HEC is expressed by several suffixes and one reduplication-based pattern. The suffix ***-uw-a** is found in Gedeo, Sidaama, and Hadiyya; ***-ey-e** is found in Gedeo, Sidaama, and Burji; and ***C_i~C_i-a** is found in Gedeo, Sidaama, and Kambaata. Burji and Kambaata instead commonly use the plurative suffixes **-na** and **-aakkata**, respectively.

Reduplication *C_i~C_i-a

Reduplication of the stem-final single consonant and addition of **-a** as a plurative formation is found in Gedeo (Eyob Kelemework, 2015, p. 107), Sidaama (Hudson, 1976, p. 252), and Kambaata (Treis, 2008, p. 135).[149] Hudson describes it as the most common form of plurative marking in Sidaama.[150] The following are examples of the reduplication-based plurative (***C_i~C_i-a**) alongside the corresponding unmarked (or singulative) base forms:

		Plurative	**Unmarked/sgv**
PHEC	***C_i~C_i-a**		
Gedeo	**n + a → n~na**	*man~n-a* 'men'	*man-jo* 'man / person'
	b + a → b~ba	*k'ub~b-a* 'fingers'	*k'ub-icco* 'finger'
Sid	**n + a → n~na**	*man~n-a* 'men'	*man-co* 'man'
	b + a → b~ba	*hogob~b-a* 'loads'	*hogob-o* 'load'
Kam	**n + a → n~na**	*min~n-a-ta* 'houses'	*min-i* 'house'
	s + a → s~sa	*boos~s-a-ta* 'water-carrying pots'	*boos-u* 'water-carrying pot'

[149] The **-a-** is described as a case marker in Kambaata followed by the feminine gender suffix **-ta** (Treis, 2008, p. 135).
[150] An alternative plurative suffix added to the same Sidaama stems is **-e** without geminating the non-geminate stem final or added to stems ending in a geminate.

Comparable reduplication-based pluratives are also attested elsewhere in Cushitic (see Zaborski, 1986; Ongaye Oda, 2013; Hayward, 1981; Wondwosen Tesfaye, 2020; Green, 2021; Tosco, 1991; Hetzron, 1976). On that basis, ***C_i~C_i-a** is plausibly a retention from PHEC; within HEC, however, the direct evidence currently comes from only three languages.

Plurative *-uw-a

The cognates of the plurative suffix ***-uw-a** are found in three HEC languages: Sidaama **-uwa** (Hudson, 1976, p. 252; Indrias Xa'miso et al., 2007, p. 635), Hadiyya **-uww-a** (Tadesse Sibamo, 2015, pp. 50, 84, 295), and Gedeo **-uwwa** (Eyob Kelemework, 2015, p. 102). The following examples illustrate the plurative suffix ***-uw-a**, showing the plurative form alongside the corresponding unmarked (or singulative) base:

		Plurative	**Unmarked/sgv**
PHEC	***C_iC_i-uw-a**		
Ged	**rr** + **uw-a** → **rr-uww-a**	*dubarr-uwwa* 'girl'	*dubarra* 'girl'
	n + **uw-a** → **n~n-uww-a**	*man~n-uwwa* 'persons'	*man-jo* 'person'
	nn + **uw-a** → **nn-uww-a**	*ann-uwwa* 'fathers'	*anna* 'father'
Sid	**nn** + **uw-a** → **nn-uw-a**	*ʔann-uwa* 'fathers'	*ʔanna* 'father'
	k'k' + **uw-a** → **k'k'-uw-a**	*sik'k'-uwa*, *sik'k'-ubba* 'stick' (pl)	*sik'k'o* 'stick'
Had	**l** + **uw-a** → **l~l-uww-a**	*k'ool~l-úwwa* 'clevers'	*k'oóra* 'clever'
	b + **uw-a** → **b~b-uww-a**	*ʔallab~b-úww-a* 'tongues'	*ʔallábo* 'tongue'

Across the examples, ***-uw-a** frequently co-occurs with stem-final consonant reduplication or glide gemination. Otherwise, it is attached to stems ending in a geminate, a cluster, or **ʔ**/**h**, e.g., Hadiyya *afoʔo*, *afoʔ-uwwa* (plv) 'hole' (Ritter, 2007, pp. 3–4). The next paragraph summarizes these distributional tendencies and exceptions.

Hudson (1976, p. 252) reconstructs ***-uba** (rather than ***-uwa**) largely on the basis of Sidaama, where intervocalic **b**, **β**, and **w** are allophonic (see §2.1.3 (PHEC *b)). He argues that stems with **b** encouraged the retention of the **b** in the suffix **-uba** (e.g., *buuda* 'horn', *buudubba* (pl)), which otherwise becomes **-uwa**. However, the distribution of Sidaama **-ubba**/**-uwa** is not limited to stems containing a labial **b**: for example, *sik'k'-uwa* and *sik'k'-ubba* 'stick' (plv) co-occur with stems lacking **b**, similarly, *ille* → *illubbà* and *illuwà* (plv) 'eye'. This weakens a strictly labial-conditioning account and leaves ***-uwa** as the more

economical comparative base and the variant with **b** as part of the allophonic alternations.

Intervocalic glide gemination may reflect a general strategy to stabilize a weak intervocalic glide, either via epenthesis and subsequent fusion (e.g., ***w** → ***w-y-** → **ww** /**V_V**) or via direct gemination. A parallel pathway is conceivable for ***-ey-e** (***y** → ***y-y-** → **yy** /**V_V**) on the assumption that the relevant glide behaves similarly in intervocalic position (see §2.1.4 (PHEC *w)). In Hadiyya and Gedeo, gemination plausibly contributes to the preservation of these sequences; in Sidaama, the intervocalic **b ~ β ~ w** alternation may instead allow the glide outcome without gemination.

Plurative ***-uw-a** is supported by cognates in Sidaama, Gedeo, and Hadiyya. It has no parallels in Kambaata or Burji. However, the plurative/collective ***-ey-e** is found in Burji, Sidaama, and Gedeo with no parallels in Kambaata (or Hadiyya).

Collective *-ey-e (f)

The cognates of the collective/plurative ***-ey-e** are found on adjectives and some nouns in Gedeo (Eyob Kelemework, 2015, p. 223), Sidaama (Kramer & Anbessa Teferra, 2020, p. 11), and Burji (Tesfaye Baye, 2015, p. 93). The following are examples of ***-ey-e** in Gedeo, Sidaama, and Burji, with the plurative/collective form next to the corresponding unmarked (or differently marked) base:[151]

		Plurative	**Unmarked/sgv**
Ged	**-eeye**	*c'im-eeye* 'old, experienced people'	*c'ima* 'old, experienced'
		korb-eeyye 'goats'	*korb-eessa* 'he-goat, billy'
Sid	**-e(e)yye**	*c'im-eeyye* 'arbitrators'	*c'im-eessa* (m), *c'im-eette* (f) 'arbitrator'
		hiyy-eyye 'wolf-like animal'	*hiyy-eyy-icco* 'wolf-like animal'
Burji	**-aayyee**/ **-aayyaa**	*c'im-aayyeé* 'elders'	*c'im-eyʃi̯* (m), *c'im-eyteé* (f) 'elder'
		ʔarʔ-eyyaá 'oxen'	*ʔarʔ-éyi̯* 'ox'

[151] Tesfaye Baye (2015, p. 93) glosses the Burji suffix as **-yyaa**, **-yyee**, but the comparative evidence suggests **-aayyaa**, **-aayyee** (cf. that *ʔarʔ-eyyaá* 'oxen' seems to be a concatenation of two suffixes given the existence of **ey** in the unmarked *ʔarʔ-éyi̯* 'ox').

The attested reflexes suggest that ***-ey-e** participates in gender/collective morphology. In dictionary sources, reflexes of ***-ey-e** are commonly described as feminine and/or collective in Gedeo (Wedekind, 2008), Sidaama (Gaspirini, 1983), and Burji (Roba Dame & Wedekind, 2008). Unlike plurative **-uw(w)-a**, these reflexes do not appear to be restricted to a particular suffixation environment.

The collective suffix ***-ey-e** is not found in Hadiyya or Kambaata. However, in Hadiyya, "[w]ith the exception of *tómmo* 'ten', *sáde* 'thirty', and *sóre* 'forty', the building morpheme for other morphologically complex tens (twenty, fifty, sixty, seventy, eighty and ninety) is [-*yye*] is (Lit:-ty)", e.g., *lamíyye* 'twenty', *ʔontáyye* 'fifty', *loháyye* 'sixty' (Tadesse Sibamo, 2015, p. 151; see comparative data in Treis, 2008, p. 305).

Both ***-uw-a** and ***-ey-e** are candidates for reconstruction to PHEC insofar as they have comparanda elsewhere in Cushitic (Zaborski, 1986; Ongaye Oda, 2013; Wondwosen Tesfaye, 2020; Mous, 1993; Mous, 2016; Harvey, 2018).

Miscellaneous Pluratives *-n, -r, -aakkata

Plurative *-n

The common plurative **-na** Burji (Tesfaye Baye, 2015, p. 90) and **-aano** in Sidaama (Anbessa Teferra, 2000, p. 39) may continue an older plurative ***-n**. This possibility is supported by (i) fossilized ***-n** forms elsewhere in HEC and (ii) the regular presence of **-n** in plural pronoun morphology.

The following list includes candidate fossilized reflexes of an older plurative ***-n** and the potential source stem without the plurative:

[142]	Fossilized plv *-n	Gloss	[254]	Source stem	Gloss
PHEC	***tam-n-**	'ten'	←	**-**	-
Burji	**tanna**	'ten'	←	***tun~tumm-** ← ***tum-**	'fist'
PNHEC	***tom-n-**	'ten'	←	**tontomma**	'fist'
Had	**tommo**	'ten'	←	**tuntummo**	'fist'
Kam	**-**	-	**-**	**tuntummu-ta**	'fist'
Sid	**tonne**	'ten'	←	**tuntummo**	'fist'
Ged	**tomme**	'ten'	←	**(kottoomoo)**	'fist'

PHEC [142] **tam-n-* 'ten' may reflect an older ***-n** formation related to the stem [254] **tun~tumm-* 'fist'. One possibility is that **tam-n-* represents a non-

reduplicated variant of the same stem complex; another is that [363] **omt-* 'five' is not derived directly from **tum-* 'fist', but via an older **tom~t-*, in which case **tam-n-* could reflect a plurative derivation within this set. The internal directionality remains uncertain.

Plurative -r

The trill **-r** is found as a plurative marker in three HEC languages: Gedeo has **-ra** (Wedekind, 2008, p. 2) or **-aro** (Eyob Kelemework, 2015, p. 223); Sidaama has in the genitive and relative formations a collective suffix **-r** (relative), **-re** (gen.acc), and **-ri** (gen.nom) (Anbessa Teferra, 2014, p. 42); and Kambaata has a nominalizing plural enclitic internally reconstructed as ***ra** (Treis, 2008, p. 362). In the rest of Cushitic, Kießling (1994, p. 53) has reconstructed a PWR collective suffix ***-arú**, e.g.,**maʕa-arú* 'small red beans'.

Plurative -k

The plurative **-aakk-a-ta**, e.g., *yann-a*, *yann-aakk-a-ta* 'proverbs' (plv) is unique within Kambaata and its related languages. This plurative formation consists of the morpheme **-aakk**, the case vowel **-a**, and the feminine gender suffix **-ta** (Treis, 2008, p. 136). The ***-k-** and ***-t-** components of this Kambaata plurative form might have parallels in one EC language. Boni has a plurative suffix **-tə** and a preceding fused or retained **k**, e.g., Boni *míŋ*, *mín* 'house' has pluratives in Boni Jara *múŋ*, *múŋ-tə* (plv) and Boni Kilii *míŋ*, *míg*, *mig-oo-tə* (plv) 'houses'. The final ***n** velarization to **ŋ** is due to assimilation to or fusion with a following **k**, which replaced the final **n** and became **g** in some examples, hence, a plurative **min-k-ta*. However, the **k** could also be posited for the unmarked variant *míŋ*. Boni Kilii *mig-oo-tə* (plv) 'houses' is comparable with the Sidaama plurative suffix *-oota* as in *man-co* (sgv), *mann-oota* (plv) 'man'.

4.1.2 Deverbative

In addition to its singulative function, the suffix complex ***-ʃ** (with an affricated allomorph ***-c**) also derives deverbal nominals across HEC. As in §4.1.1.1, the comparative evidence supports a shared PHEC inventory (***-ʃ, *ʃ-ʃ**, ***-i-ʃʃ, *-c**), but in deverbatives the choice among these surface shapes is often conditioned by preceding concatenated suffixes. The following is a summary

of the reflexes of the PHEC deverbative allomorph set across the HEC languages[152]

	PHEC	**PHEC allomorphs**	**HEC languages allomorphs**
PHEC	***-ʃ →**	***-ʃ**, ***ʃ-ʃ**, ***-i-ʃʃ**, ***-c**	
Had	……		**-anca**, **-ca**, **-i-ʃʃa**
Kam	……		**-c-a(-ta)**, **-iʃʃ-a(-ta)**
Sid	……		**-aan-co, -an-ʃo**
Ged	……		**-an-jo**, **-acco**
Burji	……		**-aancoo**, **-issoo**

Taken together, the reflexes motivate a single deverbative morpheme with multiple phonologically and morphologically conditioned realizations, rather than independent deverbative suffixes. The Gedeo form **an-jo** is compatible with the same voicing pattern discussed for the singulative in §3.1.1.5 (Affricated **j**), i.e., a voiced realization of the affricated allomorph in a restricted morphological/phonological environment, rather than a distinct deverbative suffix. Gedeo has no allomorph of post-alveolar ***ʃ**, and Burji has an assimilated allomorph **ss**. Unlike the singulative, many allomorphs of the deverbative are conditioned by a preceding concatenated suffix. Hence, the allomorph ***-iʃʃ** is mainly found on stems with a suffix **-is**, so it is explainable as **-is+ʃ/c → -iʃʃ**. Although Burji can change **ʃ** into **s** unconditionally, a concatenated **-is** likely to have assimilated the deverbative **ʃ/c** fully into **ss** in **-issoó**. The suffixation conditions of the remaining PHEC allomorphs remain the same as those of the singulative.

The deverbative functions as a word-class changing suffix to derive nominals from (verbalized) stems (§4.2.5). The following are some examples:[153]

[152] Data sources and discussion: Kambaata (Treis, 2008, pp. 158–160), Sidaama (unsegmented in the source, which is Kawachi, 2007, pp. 317–318), Gedeo (Eyob Kelemework, 2015, p. 128), Burji (Tesfaye Baye, 2015, pp. 131–132), and Hadiyya (own interpretation based on data from Ritter (2007) and Tadesse Sibamo (2015, pp. 40–41)).

[153] Data sources and further discussion: Sidaama (Yri & Pepper, 2019; Kawachi, 2007, pp. 314–317; Indrias Xa'miso et al., 2007, p. 287), Gedeo (Eyob Kelemework, 2015, p. 128), Kambaata (Treis, 2008, pp. 158–160), Burji (Tesfaye Baye, 2015, pp. 127–132; Getahun Addo, 2017, p. 593), Hadiyya (Ritter, 2007).

	Verb	Gloss	Derived Noun	Gloss
Kam	*k'al-am-*	'be born'	*k'al-an-c-a*	'generation, birth'
	kicce?-	'be(come) sad'	*kiccecc-a*	'sadness'
	hut'-is-	'make shiver'	*hut'-iʃʃ-a*	'malaria'
Had	*k'ar-am-*	'be born'	*k'ar-an-ca*	'being born'
	kicce?-	'be sad'	*kiccecca*	'sadness'
	hut'i-is-	'cause to shake'	*hut'i-iʃʃa*	'malaria'
Sid	*had'-am-*	'to walk about'	*had'-án-co*	'wanderer'
	gan-am-	'to hit, beat, fight'	*gan-an-ʃo*	'fighting'
Ged	*d'ib-em-*	'be sick'	*d'ib-em-an-jo*	'patient'
	moor-	'to steal'	*mood'-acco*	'thief'[154]
Burji	*k'un-ʔ-*	'pound'	*k'un-aancoo*	'grinding stone'
	ganʔ-	'grow'	*gann-issoo*	'growth'

Further examples with the post-alveolar **-ʃ** reflexes are discussed in §3.1.1.5 and §3.1.2.1.

The allomorph **-anc** is attested frequently across HEC. As Treis (2008, p. 160) accurately assumes, the frequent collocation of **-am** and **-c** likely constitutes a complex derivational device. Based on the semantics of a number of the derived nouns, they seem to be derived from a verbalized stem (§4.2.5), namely with suffixes as **-am**, **-ʔ**, **-is** + **-ʃ**.[155] Synchronically, Gedeo has a verbalized stem *t'uut'-em-* 'be sucked' from which the nominal *t'uut'-em-acco* 'something sucked' is derived (Eyob Kelemework, 2015, p. 132).

Yri & Pepper (2019) note that Sidaama words with such nominalizers may not have a stem without the derivational suffix **-co**, e.g., *fúr-co* 'door'. This is true for many derived lexemes. Therefore, Sidaama *fúr-co* 'door' was originally derived from the PHEC verb [143] **fur y* 'to open', even if the verbal stem is not attested synchronically (cf. Treis, 2008, p. 133).[156]

Finally, despite the formal resemblance, one should not expect the derivational deverbative suffix to fulfill a singular function besides the nominalizing function.

In addition to **-c** and **-ʃ**, three HEC languages have a deverbative with **-t**: Sidaama and Gedeo **-atto**, and Burji **-too** (feminine agentive/experiencer).

[154] The deverbative *mood'-acco* 'thief' is derived via an intermediate verb **moor-ʔ-*.
[155] Various deverbative markers are found attached to PHEC roots, such as [198] **tum-am-ʃ-*, **tum-t-iʃʃ-* 'blacksmith' from the verb [48] **tum-* 'to pound'. The reconstructed deverbative stems in this study do not all include the lexicalized denominative morphemes, but they should be considered upon further research.
[156] See Gedeo *farriyya* 'be clear, open' (Wedekind, 2008, p. 146).

The following examples from Gedeo and Burji illustrate this derivation, contrasting the base verb with the agentive or instrumental nominal:

	Verb	**Gloss**	**Derived Noun**	**Gloss**
Gedeo	*k'oc'-*	'to pick'	*k'oc'-atto*	'one who picks'
Burji	*k'un-*	'pound'	*k'un-too*	'pounder'
	ʔuum-	'create'	*ʔuun-too*	'creator'

While the derivational function is clear, the regular HEC correspondences for these specific variants require further investigation (see §3.1.1.5).

4.1.3 Problematic Extension

A small set of nominal stems shows irregular variation between forms with a glottal stop or fused glottalized consonants and parallel forms without glottalization. The conditioning for this glottalization is opaque, though the evidence permits several potential sources: (i) remnants of an archaic extension (possibly ***ʔ**), (ii) irregular or analogical glottalization, (iii) obstruent (glottalization) harmony (fn. 36), or (iv) multiple derivational pathways.

Comparative set [145] exemplifies this divergence, showing a split between Burji/Gedeo (glottalized) and the remaining HEC languages (plain).

[145]	**Glottal stop**	**No glottal stop**	Gloss
Burji	**ʃinʔaa**	**-**	'urine'
Had	**-**	**ʃuma**	'urine'
Kam	**-**	**ʃuma-ta**	'urine'
Sid	**-**	**ʃuma**	'urine'
Ged	**ʃiʔna**	**-**	'urine'

The glottal pattern in [145] suggests a morphological rather than purely phonological origin. Specifically, the glottalized variants likely reflect a cycle of cross-class derivation (noun → verb → noun). For instance, [145] may result from compounding with the support verb *ʔ* 'to do' (§§4.2.3, 4.2.2.1, 4.1.4.1), yielding a PHEC form like [170] **sum ʔ* 'urinate', which subsequently lexicalized as a noun in specific daughter languages.

A similar split appears in [163], where the glottal stop is retained in Burji and Hadiyya but absent in Kambaata and Sidaama, and a cross-class derivation with ***ʔ** 'to do' is not feasible.

[163]	**Glottal stop**	**No glottal stop**	**Gloss**
Burji	**(h)arʔeyi, aʔré**	**-**	‘ox’, ‘bull’
Had	**woʔla**	**-**	‘calf (of young cow)’
Kam	**-**	**worjaamu**	‘young bull’
Sid	**-**	**wor-aamo**	‘ox, bull’

Two further comparative sets, [46] and [180], illustrate the extent of this variation, highlighting both cross-language mismatches and language-internal doublets.

[46]	**Glottalized (C’)C’**	**Non-glottalized**	**Gloss**
Burji	**irk’a**	**-**	‘tooth’
Had	**ink’e**	**-**	‘tooth’
Kam	**ink’u-ta**	**-**	‘tooth’
Sid	**-**	**hinko**	‘tooth’

The complex glottal reflexes for this Common EC stem are discussed in §3.3.1.3 and §3.4.3.

[180]	**Glottalized (C’)C’**	**Non-glottalized**	**Gloss**
Burji	**-**	**utubaa**	‘pole, center-pole of house’
Had	**-**	**utuba**	‘pole, center-pole of house’
Kam	**-**	**utubu**	‘pole, center-pole of house’
Sid	**-**	**ʔutuba**	‘private property’
Ged	**utup’a**	**utuba**	‘support, e.g., forked stick’, ‘pole, center-pole of house’

Gedeo shows both glottalized *utup’a* and non-glottalized *utuba* reflexes for [180]. As both function as nominal stems, the glottalization likely stems from the derivational history, though Oromo *utubaa* ‘center-pole’ offers a potential contact source for the plain variant.

Two further sets illustrate comparable irregularities in glottalization.

> Had [81] *kobeʔe*, Sid *kop’é*, Ged *kop’e* ‘shoes’ (see Oro *kop’-ee*, *kob-ee* ‘sandal’);
>
> Kam [544] *c’up’i-ta*, Ged *c’uup’e* ‘dagger’ (cf. Amharic *c’ube* and Oro *c’uubee* ‘dagger’)

Finally, certain irregularities resolve as clear phonological repairs. The intervocalic glottal stop in Hadiyya [81] *kobeʔe* ‘shoes’ and [544] *c’ubeʔe* ‘dagger’, for example, functions as an epenthetic insertion to resolve vowel

hiatus (§3.3.4.1), distinguishing it from the opaque morphological glottalization discussed above.[157] The glottalized Kambaata and Gedeo **p'** in [544] is likely due to obstruent (glottalization) harmony (fn. 36), similar to the following example [49] (with full assimilation in Hadiyya):

> Had [49] *k'uunk'a*, Kam *k'uup'p'-iccu-ta*, Sid *k'uup'p'e*, Ged *k'uup'e*; *k'uup'p'e* 'egg(s)'.

Alternatively, the fused glottalized **p'p'** in [49] could be derived via the verbalizer **ʔ** (§4.2.5.1); for instance, Kambaata *k'uup'p'-iccu-ta* 'egg' might be derived from the nominal *k'ububb-eema* 'round, circular, ball-shaped' via the regular verbalizer **k'uup'p'-* 'be(come) circular' then used as a noun *k'uup'p'-iccu-ta* 'egg' (see the current verbalizer *k'ububb y-* 'be(come) circular)'.

In addition to the derivation explanation, the examples certainly include loan words; for instance, in [81], Sidaama *kop'é*, Gedeo *kop'e* with irregular single **p'** instead of the expected geminated glottalized **p'p'** (§4.1.3) might be a loan from Oromo, where both *kop'-ee* and *kob-ee* 'sandal' exist. The irregular single **p'** is also noticeable in [544] and [180]. However, the problematic examples cannot all be loans, for instance, [46] is attested throughout Cushitic usually as *ilk-* 'tooth'.[158]

4.1.4 Genitive

The genitive construction expresses possession or relational attributes between nouns in HEC. Across the five languages, genitive marking ranges from the conservative retention of a glottal consonant (Hadiyya) through vowel and suprafix systems (Sidaama, Kambaata) to morphologically complex clitic constructions (Burji, Gedeo). This section presents the comparative evidence and examines each genitive type: the glottal marker, vowel/suprafix systems, and the gender-clitic compounds (§4.1.4.1); gives an overview of the genitive pronouns (§4.1.4.2) and genitive components (§4.1.4.3); and concludes with a proposed historical scenario (§4.1.4.4). The reconstruction argues for a PHEC genitive suffix ***-ʔ**, whose reflexes are obscured by subsequent phonological processes but can be traced through Hadiyya glottal evidence,

[157] Similarly, **-ʔ-** epenthesis in Hadiyya [70] *afoʔo* 'hole' and Gedeo *afaʔo* 'mouth, language, opening', and **-y-** in Burji *afeyi* 'mouth, language' (§3.4.1.2).
[158] The Ts'amakko parallel is *ʔilǧe* 'tooth'.

Gedeo glottalized clitics, and compensatory doubling in Kambaata and Sidaama.

4.1.4.1 Genitive Markers

Highland East Cushitic languages use different forms of genitive morphemes depending on gender, case, and nominal type. The following is a summary of the principal genitive exponents attested on common nouns, proper nouns, and pronouns in each language. The **-**◌́ indicates a high-pitch suprafix without a segmental vowel, and **-C'C'** represents geminated glottalized consonants in Hadiyya.

	Common Nouns	**Proper Nouns**	**Pronouns**
Burji	**-ng**, **-nk** (m), **-nd**, **-nt** (f)	**-kk** (m), **-tt** (f)	**-kk** (m), **-tt** (f)
Ged	**-nk**, **-k** (m), **-nt'**, **-(t')t'** (f)	**-k** (m), **-t't'** (f)	**-k** (m), **-t't'** (f)
Sid	**-ú**, **-í** (m), **-te** (f), **-**◌́	**-i**, **-ø**	**-í** (m), **-é** (f)
Kam	**-V́(V)** (excluding **u**)	**-V́** (excluding **u**)	**-í** (m), **-é** (f)
Had	**-ʔ**, **-C'C'**, word order	**-**	**-t-ʔ** (m)

The markers show considerable formal diversity, and each is analyzed in the subsequent subsections. The reconstructive challenge is to identify the common source underlying this surface variation.

The ordering of the following subsections proceeds from what is hypothesized to be the most conservative trace of an ancient genitive marker (the glottal stop in Hadiyya), through intermediate stages involving phonological loss and change (vowel/suprafix systems in Kambaata/Sidaama), to the most morphologically complex and potentially innovative systems involving fused or layered clitics (Burji/Gedeo). This structure aims to build the case for the historical hypothesis presented at the end by showing the different outcomes across the language group.

Genitive *-ʔ

Tracing the history of HEC genitives, the reflexes involving a glottal stop, found predominantly in Hadiyya and also able to be deduced in Gedeo, potentially represent the remnants of an older genitive marking strategy, possibly involving a PHEC ***-ʔ** marker (cf. reflexive ***-ʔ** in §4.2.3.1). Examining

these forms provides a baseline for understanding subsequent innovations and diversifications across the language group.[159]

		Example	Gloss
Had	**-ʔ**	*maára* 'meat' + *mine* 'house' → *maaʔli míne*[160]	'butchery'
	-p'p'	*kitaaba* 'book' + *mine* 'house' *kitaap'p'i mine*	'bookstore'
	-k'k'	*elektiriiksiiteʔe* 'electricity' + *amat't'a* 'volume' → *elektriik'k'i ammat't'a*	'voltage'
	-t't'	*diinate* 'money' + *k'at'ooma* 'quantity' → *diinat't'i k'at'ooma*	'quantity of money'
Ged	**-t't'a** (f.acc)	*jissot't'a beltu*	'Jisso's girl'
	-t't'i (f.nom)	*manjiccotit't'i belto*	'the daughter of the man/woman'
	-nt'a (f.acc)	*warfint'a isso*	'dog's teeth'

Hudson (1976, p. 254) and Tadesse Sibamo (2015, p. 60) mention a glottalized (stem-)final sonorant when a noun is genitivized in Hadiyya. However, neither recognizes the glottal stop as a genitive extension or identify the source of such glottalization. The above is based on my interpretation of the available data from those sources and Ritter (2007), per comparable behavior of the glottal stop in such contexts in HEC (§3.3).

Thus, the glottal stop **ʔ** in the above examples appears in compounds and genitive phrases, even when absent from the base forms in Hadiyya. This suggests a morphophonological process, possibly triggered by a genitive marker **-ʔ**. In Gedeo, the glottalized **-t't'** (see the pronoun *iset't'e* 'her' (Eyob Kelemework, 2015, p. 267)) could also reflect a glottalized feminine marker **-t** resulting from fusion with a following glottal stop. Like Hadiyya, it might derive from a similar genitive **-ʔ**.

Treis (2008, p. 189) identifies a glottal element in the genitive and ablative forms as an unknown potential extension, e.g., *ber-é* (acc) 'yesterday', *ber-é ~ ber-e-ʔée* (gen), *ber-éecc ~ ber-e-ʔéecc* (ablative) (compare with the dative *ber-ée*) and *t'on-íta* (acc) 'before, recently', *t'on-é ~ t'on-e-ʔée* (gen). Later, she describes this glottal extension (at least for the genitive) as reminiscent of

[159] Data sources: Gedeo (Hudson, 1976, p. 254; Eyob Kelemework, 2015, pp. 101, 236); Hadiyya (Tadesse Sibamo, 2015, pp. 38, 60; Ritter, 2007, pp. 75, 82).
[160] Tadesse Sibamo (2015, p. 38) transcribes it as one word: *maaʔlimíne* 'butchery'. However, it is most likely a phrase, as consistently transcribed in the data from Ritter (2007), e.g., *laʔl mine* 'cattle area (in house)'.

the associative morpheme. A comparable glottal stop has been described as inserted before the derivational morpheme, e.g., *ceemm-í-ta* (acc), *ceemm-é* (gen) 'laziness', *ceemm-e?-aam-ú* (m.adj), *ceemm-e?-aam-í-ta* (f.adj) 'lazy' (Treis, 2008, p. 275). The latter analysis is the one adopted for all these glottal stops and has historical support in the form of epenthesis between three consonants (§3.3.4).

Genitive Vowel and Suprafix

The genitive systems in Sidaama and Kambaata use vowels and suprasegmental high pitch rather than the common overt consonantal markers. This likely reflects a distinct route of development characterized by the loss of an earlier consonantal element, potentially the ***-?** genitive or a subsequent clitic, followed by compensatory phonological adjustments such as vowel coalescence or accentuation, leading to the synchronic vowel and suprafix markers. The glottal stop is a common source of a high tone in various languages (Hombert et al., 1979, pp. 49–51).[161]

		Example	Gloss
Sid	**-u** (m Unmod)	*beett-ú faraʃʃ-i*	'The boy's horse'
	-i (m Mod)	*dangur-í beett-í faraʃʃ-i*	'Dangura's boy's horse'
	-i (m proper N)	*dangur-í faraʃʃ-i*	'Dangura's horse'
	-ó (f Mod)	*dangur-í beettó faraʃʃ-i*	'Dangura's girl's horse'
	-ó (f proper N)	*bulé faraʃʃ-i*	'Bule's horse'
Kam	**-á(a)** (f, m)	*roʃʃ-á (f) min-í*	'school'
		had-á (f) baad-ú	'abroad'
	-é(e) (f, m)	*Ludag-é (f) ball-úta*	'Ludage's mother-in-law'
	-ó(o) (f, m)	*az-ó (f) saʔ-á*	'dairy cow'
	-í(i) (m)	*zaneecc-í (m) girk'-á*	'tusk of the / an elephant'

Sidaama employs vowels similar to those of the nominative, i.e., **-u** and **-i**, as a genitive marker of Unmodified (Unmod) and Modified (Mod) common masculine nouns, respectively.[162] The genitive **-i** is also attached to some

[161] Data sources: Sidaama (Kawachi, 2007, pp. 357–359), Kambaata (Treis, 2008, pp. 391, 383, 119). Some of these constructions are taken from longer contexts in the original sources, resulting in them potentially yielding different meanings or being difficult to translate when used in isolation. Therefore, the page numbers are provided for reference to the full context.

[162] For Kawachi (2007, pp. 247–248), "…"Modified" and "Unmodified" are capitalized because these terms are used here to refer to a morphosyntactic distinction specific to common nouns in Sidaama... Modified common nouns include not only those that have genitive NP dependents or adnominal dependents, but also those accompanied by the possessive pronominal suffix. Unmodified common nouns are those without any of

proper nouns, e.g., *dangur-í*. Additionally, the genitive vowels receive an accent as a genitive suprafix marker (Kawachi, 2007, pp. 356–357). The feminine common nouns, on the other hand, receive **-te** (Unmod), e.g., *beettó-te faraʃʃ-i* 'the girl's horse', and **-ø** (Modified and proper). When the feminine genitive suffix is **-ø**, the high pitch on the final vowel of the stem is analyzed as the sole genitive marker (Kawachi, 2007, p. 359) (e.g., *beettó*, *bulé* in the examples above).

Except **u**, all other vowels (**a**, **e**, **o**, and **i**) are attested in Kambaata marking the genitive, both as moraically single and double. In the genitive, they also receive a suprafix as an accent **-V́**, **-V́V** (Treis, 2008, p. 103). However, in contrast to Sidaama (Kawachi, 2007, p. 360), the accent on genitive nouns is sensitive to the length of a word in Kambaata; i.e., when an accent-neutral suffix extends the word, the accent shifts as far to the right as possible, e.g., the genitive *bit'-í* 'of a stick' → *bit'-i-ʔné* 'of your (plural) stick', *bit'-i-kkiʔné* 'of your (honorific) stick' (see also *gujj-í* → *gujj-i-sí* 'of his back', *ball-ó* → *ball-o-sé* 'of her mother-in-law', *hill-ée* → *hill-ee-ssá* 'of their intestine' (Treis, 2008, pp. 55, 107)).

Here is a summary of the similarities and differences between Sidaama and Kambaata vowel and suprafix genitives (**x** means attested, **-** indicates unattested):

	Short V	Double V	V(V) + Suprafix	Suprafix only	Suprafix shift
Sidaama	**x**	**-**	**x**	**x**	**-**
Kambaata	**x**	**x**	**x**	**-**	**x**

The two systems are too similar to be analyzed differently. The historical scenarios proposed below in §4.1.4.4 provide an alternative explanation that applies to both.

The vowels attested following the genitive clitics **=k-/=h-** and **=t-** below are usually associated with certain cases or predicative use.

these "Modifiers" (those neither with a dependent or with the possessive pronominal suffix)".

		=k-/=h-	**=t-**
Burji	Nominative	**-u**	**-i**
	Accusative	**-a**	**-a**
Ged	Nominative	**-i**	**-i**
	Accusative	**-a**	**-a**
	Predicative	**-e**	**-e**
Sid	Predicative clitic	**-o**	**-e**
	Clause final	**-**	**-i**
Kam	Clitic	**-a**	**-a**

The vowels following the **=k-** and **=t-** in Burji and Gedeo seem to be determined by the nominative-accusative distinction. Eyob Kelemework (2015, p. 338) notes that "[o]ne important aspect of Gedeo genitive case is that the form of genitive clitic shows variation in the nominative and accusative case. In the nominative case, the nominative case marking suffix -i would be suffixed, but -a is used in the accusative", e.g., the nominative *manjiccinki* becomes *manjiccinka* 'of the man' in the accusative. The same could be inferred from the Burji data. The accusative takes **-a**, but there are further vowel distinctions depending on the gender of the nominative: **-i** for the feminine and **-u** for the masculine.

The following sections distinguish between "Genitive Clitics **=(k)k**/**=(t)t**" and "Gender Clitics **=k**/**=t**". The former refers to the often phonologically fused or assimilated forms (e.g., geminates) integral to the main genitive constructions in languages like Burji and Gedeo. The latter term, "Gender Clitics", is used more broadly to discuss the **k**/**t** elements when they appear in other contexts or languages (e.g., Kambaata linkers/copulas, Sidaama predicative clitics) where their function as more independent agreement markers, linkers, or copula elements is clearer. This distinction helps highlight the potential functional origins and broader distribution of the **k**/**t** elements that became grammaticalized within the specific genitive structures of Burji and Gedeo.

Genitive Clitics =(k)k and =(t)t

Burji and Gedeo represent the most morphologically elaborate genitive systems in HEC. Both languages employ clitics **=k** (m) and **=t** (f), which interact with the genitive marker ***-ʔ** to yield geminated (**=kk**, **=tt**) or fused (**=t't'**) forms on proper nouns and pronouns, and epenthetic-nasal forms (**n=k**, **n=t**) on common nouns due to degemination. These processes suggest a stage involving the integration or fusion of the gender clitic with a preceding element, potentially the old genitive marker ***-ʔ**, as outlined in the final

hypothesis (§4.1.4.4). Sidaama retains a reflex only in the feminine **-te**. This paradigmatic complexity distinguishes the southern HEC languages from the simpler systems of Hadiyya and Kambaata.[163]

		Example	Gloss
Burji	**-kku** (m.nom)	*bariteékku ʔaábbu̱*	'Barite's father' (subj)
	-kka (m.acc)	*bariteékka̱ ʔaábboó*	'Barite's father' (obj)
	n=gu (m.nom)	*halamíngu̱ wóccu̱*	'(the) boy's dog' (subj)
	n=ga (m.acc)	*halamínga̱ woccoó*	'(the) boy's dog' (obj)
	-cci (f.nom)	*baritéccị̱ ʔaáyyị̱*	'Barite's mother' (subj)
	-tta (f.acc)	*baritéttа̱ ʔaáyyeé*	'Barite's mother' (obj)
	n=ci (f.nom)	*halamíncị̱ hárrị̱*	'(the) boy's donkey' (subj)
	n=da (f.acc)	*halamínda̱ harreé*	'(the) boy's donkey' (obj)
Ged	**=ki** (m.nom)	*jissoki belti*	'Jisso's son' (subj)
	n=ki (m.nom)	*manninki min-i*	'the people's houses'
	=ka (m.acc)	*belto jissoka*	'Jisso's son' (obj)
	n=ka (m.acc)	*ayyaaninka lukkicco*	'a hen for a holiday'
	=ke (m.pred)	*weesetike ficcani*	'weese-plant's farm'
	n=ke (m.pred)	*bisinke gogani*	'the skin of a body'
	=t'i (f.nom)	*jissot'i belti*	'Jisso's daughter' (subj)
	n=t'i (f.nom)	*lamaʔint'i beltanna*	'the young of a rat' (subj)
	=t't'a (f.acc)	*belto jissot't'a*	'Jisso's daughter' (obj)
	n=t'a (f.acc)	*hajyyomaint'a galca*	'position of power' (obj)
	=(t')t'e (f.pred)	*hak'iccit't'e lame dame-ʔnni*	'the two wings of a tree' (pred)
	n=t'e (f.pred)	*hak'iccint'e dame*	'tree's wing' (pred)
Sid	**-te** (f)	*beettó-te faraʃʃ-i*	'the girl's horse'

Burji and Gedeo morphemes with **-n-** are found on common nouns (§3.4.3) while the geminate variants are mostly found on proper nouns and pronouns. Burji morphemes **n=gu** (m.nom), **n=ci** (f.nom), **n=ga** (m.acc), and **n=da** (f.acc)[164] are found on common nouns, whereas **-kku** (m.nom), and **-cci** (f.nom) are found on proper nouns (see §3.2.1 for palatalization and §3.1.1.1 voicing assimilation).

For Gedeo, Hudson (1976, p. 254) has **-nka** and **-nt'a** (accusative), **-nki** and **-nt'i** (nominative), e.g., *lalinka goga* 'a cow-hide' and *warʃint'a isso* 'dog's

[163] Data sources: Burji (Tesfaye Baye, 2015), Gedeo (Eyob Kelemework, 2015), Sidaama (Kawachi, 2007).

[164] Tesfaye Baye (2015) segments the Burji morphemes as suffixes: **-ngu** (m.nom), **-nci** (f.nom), **-nga** (m.acc), and **-nda** (f.acc). The above clitic-based segmentation (with epenthetic **n**) is used for ease of reading, since this matches the pattern in the rest of the HEC languages.

teeth' (see section on Nasal **n** below). Eyob Kelemework (2015, p. 270) specifies the differences as follows: **=ki, =(t)t'i** (nominative), **=ka, =(t')t'a** (accusative), and **=ke, =(t')t'e** (predicative), with a difference in the vowel quality, and **k** on the masculine and **t** on the feminine forms. Gedeo also shows a unique switch (apposition) in the order of the possessor and possessed in the object genitive constructions as discussed in §4.1.4.3, e.g., *belto jissoka* 'Jisso's son' (obj) vs. *jissoki belti* 'Jisso's son' (subj) (see Eyob Kelemework, 2015, p. 236).

Sidaama **-te** is found on Unmodified feminine common nouns. The genitive of the Modified common nouns is marked with a high-pitch suprafix **-**◌́ on the stem-final vowel only.

Gender Clitics =k and =t

While the previous section examined the heavily integrated and often phonologically altered forms of the **k**/**t** clitics in Burji and Gedeo genitives, similar elements also appear in other constructions across HEC, notably in Sidaama and Kambaata, often functioning more clearly as independent gender/agreement markers, linkers, or even copulas. Analyzing these functions provides crucial context for understanding the potential origin and functional range of the **k**/**t** elements that became integral parts of the genitive constructions in languages like Gedeo and Burji.

Sidaama uses the clitics =**h** (m) and =**t** (f) as dependent morphemes added to genitive constructions. In Kambaata, "[t]he possessive suffix … is attached after the constituent consisting of head and -ha / -ta ... An obligatory linker morpheme -n is inserted between -ha / -ta and the possessive suffix ... Predicate noun, -ha / -ta, and possessive suffix form a single phonological word" (Treis, 2008, p. 418). Hudson (1976, p. 254) describes **-ha** (**-ka**) (m) / **-ta** (f) as gender or 'article' suffixes that occur in genitive constructions with pronoun possessors in Kambaata. Note that Treis (2008, p. 418) does not identify **-ha** (**-ka**) (m) / **-ta** (f) as articles but rather as a copula or a linker.[165]

[165] Anbessa Teferra (2014, pp. 34–36) claims that the suffixes **-k** (m) and **-t** (f) are found on pronouns and common/proper nouns (as) genitive constructions in Sidaama. However, such analysis is challenged when compared to Kawachi's analysis of Sidaama.

		Example	Gloss
Sid	**=ho** (m)	*kuni mini danguri ann-i=ho*	'this house is Dangura's father's'
	-te (f)	*beettó-te faraʃʃ-i*	'the girl's horse'
	=ti (m/f)	*hakku faraʃʃo-se=ti*	'that is her horse'
Kam	**=ha** (m)	*níi báad-i=a gaʃʃáancu*	'(it is) the ruler of our country'
	-ta (f)	*aaz-éen-ta-s*	'inside of it'

The genitives of Sidaama could be classified into attributive (adnominal), e.g., *beettó-te faraʃʃ-i* 'the girl's horse', and predicative. The predicative genitive constructions receive **=te** (f) and **=ho** (m), i.e., predicative noun-phrase clitics (npc.pred) in Kawachi's (2007, p. 250) words. The clitics are suffixed to the Modified common nouns, proper nouns, and pronouns, e.g.:

kuni mini danguri ann-i=ho
Dangura.gen.m father-gen.m=npc.m.pred
'This house is Dangura's father's.'

kuni mini isí=ho
3sg.m.gen=npc.m.pred
'This house is his.'

hatti ooso dangur-í=te
Dangura-gen.m=npc.f.pred
'Those children are Dangura's'

They are also suffixed to the Unmodified common nouns preceded by the suffix **-nni**, e.g.:

kuni mini t'agisaanc-u-nni=ho
healer-gen.m-nni=npc.m.pred
'This house is the healer's'

Additionally, **=ti** occurs clause-finally for both masculine and feminine, such as on the pronoun (Kawachi, 2007, pp. 250–251, 356, 247).

hakku faraʃʃo-se=ti
horse-3sg.f.poss=npc.pred
'that is her horse'

Gedeo equivalents of predicative noun-phrase clitics in Sidaama are **=te**/=**t't'e** (f.sg/pl), **=ke** (m.sg/pl), and **=ra** (pl), which are also labeled as NPCs (Eyob Kelemework, 2015, p. 298). Gedeo NPCs are indistinguishable from the

predicative genitives **=ke** and **=(t')t'e**, hence the proposed classification of genitives in Sidaama into attributive and predicative.

However, Anbessa Teferra (2014, pp. 34–36) analyzes the predicative genitive construction in Sidaama differently. The genitive suffixes of the proper nouns, common nouns, and pronouns in such constructions include, for the nominative, **-h-u** (gen.m-nom), **-t-i** (gen.f-nom), and **-r-i** (gen.pl-nom) in the following examples:

beetto-nni-h-u
girl-perp-gen.m-nom
'the girl's' (m.subj)

beetto-nni-t-i
girl-prep-gen.f-nom
'the girl's' (f.subj)

beetto-nni-r-i
girl-prep-gen.pl-nom
'the girl's (pl.subj)'

Additionally, they include, for the absolutive (or accusative), **-ha** (gen.m.abs), **-ta** (gen.f.abs), and **-re** (gen.pl.abs) in the following examples:

beetto-nni-ha
girl-prep-gen.m.abs
'the girl's (m.obj)'

beetto-nni-ta
girl-prep-gen.f.abs
'the girl's (f.obj)'

beetto-nni-re
girl-prep-gen.pl.abs)
'the girl's (pl.obj)'

The following is a summary of the differences in analyses of Sidaama:

Feature	Kawachi (2007) Analysis	Anbessa Teferra (2014) Analysis
Core Markers	Predicate noun-phrase clitics: **=ho** (m), **=te** (f)	Genitive suffixes: **-h-u** (m), **-t-i** (f), **-r-i** (pl)
Marker Final Vowel	Part of the clitic	Case markers (nominative/absolutive)
Stem-final Vowel	Genitive marker (vowel/suprafix)	Part of the stem
Conditioning	Based on noun modification status (Modified/Unmodified)	Based on case (nominative/absolutive)
-nni Element	Suffix preceding clitic on Unmodified common nouns	Preposition preceding genitive marker
Plural Marker	Not explicitly part of this genitive system analysis	Genitive suffix **-r-** identified

In addition to the variation in the morphological analysis (Modified/Unmodified vs. nominative/absolutive), Anbessa Teferra's interpretation differs from Kawachi's in the description of the consonants and vowels, with differences even in their quality and assigned function. Kawachi interprets **=t** and **=h** as predicate clitics and the vowel following the stem or the suprafix (see the section on Genitive Vowel and Suprafix) as the genitive markers. Meanwhile, for Anbessa Teferra, **-t** and **-h** are the actual genitive markers, and the vowels are nominative/absolutive case markers. Anbessa Teferra's additional **-r** has been reported as animate **=re** (pl.acc), =**ri** (pl.nom/gen) in Sidaama (Kawachi, 2007, p. 243), and nominalizer **=ra** in Kambaata (Treis, 2008, p. 240). Therefore, the **-h-**, **-t-**, and **-r-** in Anbessa Teferra's examples might be nominalizers, e.g., 'the girl's (one(s))' (Treis, p.c.).[166]

In Kambaata, **-ha**/**-ta** are found in more complex nominal constructions involving the genitive and possessive pronouns. Treis (2008, pp. 351–352, 418) analyses them as enclitic, linker (l), and copula (cop2) depending on the nominal construction, e.g.:

locative: aaz-éen-ta-s
interior-m.loc-l-3sg.m.poss
'inside of it' (p. 195)

[166] "Headless genitive nouns are combined with the nominalizer =ra ...."
maal-i=r-a
meat-m.gen=ra-m.acc
'the ones of meat' (here: the hens that are suitable for meat production) (Treis, 2008, p. 240)

dative: hiz-éeha-n-se
sibling-m.dat-l-3sg.f.poss
'to her brother' (p. 226)

oblique: bahír-ua-n-s hiz-éeha
elder-m.obl-3sg.m.poss sibling-m.dat
'to his elder brother' (p. 226)

instrumental comitative perlative (icp): ʃiin-íin-ka-s
side-m.icp-l-3sg.m.poss
'his side' (p. 209)

ísi-ta béet-u
3sg.m.gen-f.cop2 child-f.pred
'(it) is his daughter' (p. 418)

predicate: níi báadi-a gaʃʃ-aan-c-ú
1pl.gen country-m.gen-m.cop2 rule-aan-sg-m.pred
'(he) is the ruler of our country' (p. 414)

mín-i-a-n-se-baʔa
house-m.pred-cop2-l-3sg.f.poss-neg
'(it) is not her house' (p. 352)

The examples show the lenition of **k** → **h** → **ø**, which is confirmed by Treis' (2008, p. 352, fn. 31) observation that "[t]he use of a (masculine) linker morpheme -ka after masculine nouns is not rejected as ungrammatical, but it is hardly attested in the corpus".

	masculine **-k-**	feminine **-t-**	plural **-r-**
Ged	**=ke** (npc)	**=te**/=**t'e** (npc)	**=ra** (npc)
Sid	**=ho** (npc)	**=te** (npc)	**=ri**
Kam	**-ha**	**-ta**	**=ra** (nominalizer)

In addition to the nominal functions, the terminological consensus in Gedeo (Eyob Kelemework, 2015, p. 270) and Sidaama (Kawachi, 2007, p. 250) is that **=k**/**=h** and **=t** are (en)clitics. Therefore, clitic is used here as a cover term wherever those exponents are used with the genitive. The clitic analysis is supported by the fact that, although the modifier receives the clitic, the gender of the specific clitic must agree with that of the head noun (see the subsection on Gender) at least in Gedeo and Burji. The Kambaata suffixes **-h**/**-t** are most

likely unrelated to the clitics of Gedeo and Burji and are instead similar to the common exponents found in other nominals (see §4.1.5.1), including the prominent feminine accusative **-ta**.

Offering a compatible historical perspective, Tosco (1994) proposes a specific origin for these widespread **k**/**t** elements, which he terms “Gender Markers”. He argues they derive from grammaticalized copulas used in originally left-dislocated cleft constructions (Modifier + Copula + Head). Over time, the copula, following the modifier, was reinterpreted as an agreement marker attached to the possessor but agreeing with the possessed noun. Tosco finds evidence for this copular origin in the formal similarities between these markers and overt copulas still present in several HEC languages. He further suggests these **k**/**t** elements might have initially functioned more broadly as determiners or invariant copulas before being specialized for gender marking.

Nasal n

A further layer of complexity, primarily affecting common nouns in Burji, Gedeo, and specific Sidaama constructions, involves the appearance of a nasal element **n** preceding the **k** or **t** clitic. This nasal is arguably not part of the original clitic structure but rather a later phonological insertion (epenthesis) that arose following the simplification (degemination) of earlier fused geminate forms like **kk** and **tt**, preserving a trace of intermediate geminated consonants (§3.4.3).

		Example	Gloss
Burji	**n=gu** (m.nom)	*halamíngụ wóccụ*	‘(the) boy’s dog’ (subj)
	n=ga (m.acc)	*halamínga woccoó*	‘(the) boy’s dog’ (obj)
	n=ci (f.nom)	*halamíncị hárrị*	‘(the) boy’s donkey’ (subj)
	n=da (f.acc)	*halamíndạ harreé*	‘(the) boy’s donkey’ (obj)
Ged	**n=ki** (m.nom)	*manninki min-i*	‘the people’s houses’
	n=ka (m.acc)	*ayyaaninka lukkicco*	‘a hen for a holiday’
	n=ke (m.pred)	*bisinke gogani*	‘the skin of a body’
	n=t’i (f.nom)	*lamaʔint’i beltanna*	‘the young of a rat’ (subj)
	n=t’a (f.acc)	*hajyyomaint’a galca*	‘position of power’
	n=t’e (f.pred)	*hak’iccint’e dame*	‘tree’s wing’
Sid	**nni=ho** (m)	*t’agisaancunniho*	‘(it is) the healer’s’
	nni=ti (f)	*woʃiccunniti wajjido lekka*	‘The dog’s (m) white (f) legs (f)’

Burji **-ngu** (m.nom), **-nci** (f.nom), **-nga** (m.acc), and **-nda** (f.acc) are found on common nouns, corresponding to **-kku** (m.nom), **-cci** (f.nom) on the proper nouns (see §3.2.1 for palatalization and §3.1.1.1 voicing assimilation). For

Burji, Hudson (1976, p. 254; 1981, p. 102) has voiceless **-nka**, **-nta** (which could represent a dialectal difference), and no vowel following **-kk** or **-tt** (but there are devoiced vowels in Tesfaye Baye (2015)). In Gedeo, instead of emphatic **-n** and **=t'e** / **=t'i** / **=ke** / **=ki** / **=ka** (Eyob Kelemework, 2015, p. 117), Hudson (1976, p. 254) has **-nka** and **-nt'a** (accusative), **-nki** and **-nt'i** (nominative) on the common nouns, e.g., *lalinka goga* 'a cow-hide' and *warfint'a isso* 'dog's teeth'.

The **n** before **k** and **t** is described as a linker by Hudson (1981, pp. 102, 108) in Burji and Gedeo and emphatic in Gedeo (Eyob Kelemework, 2015, pp. 116–117, 338).[167] In Sidaama, **-nni** is described as an obligatory definite suffix preceding the clitics of the Unmodified genitive noun (Kawachi, 2007, p. 384), while Anbessa Teferra (2014, p. 36) identifies it as a postposition marker.[168]

For Tosco (1994), however, the interpretation of this **n** element in Burji and Gedeo differs from the above synchronic linker or emphatic marker descriptions. He proposes that it represents a historically later determinative element, introduced specifically with common nouns after the original determinative function of the **k**/**t** markers themselves had faded. The typical absence of this **n** with inherently definite proper nouns and possibly the definite suffix **-nni** in Sidaama (Kawachi, 2007, pp. 77–78) supports this interpretation of **n** as a secondary definiteness marker. However, the analysis adopted here reinterprets this sequence: The geminates ***kk** and ***tt** result from the fusion of ***ʔ=k** and ***ʔ=t** (genitive + clitic), with **-n-** arising upon degemination to maintain the biconsonotal construction.

Overall, the genitive marker ***-ʔ** is reconstructed at the PHEC level as arguable: Direct evidence survives in Hadiyya (overt glottal stop and **C'C'** glottalization), Gedeo (glottalized **t't'** feminine), and indirectly in Kambaata and Sidaama (compensatory vowel doubling and accentuation following glottal loss). Burji and Gedeo additionally preserve the gendered clitics ***=k** (m) and ***=t** (f), which represent concatenated suffixes attached to the genitive construction. The proto-status of the combined construction ***-ʔ=k/t** is tentative, given that Hadiyya and Kambaata show no synchronic trace of the velar/dental clitics.

---

[167] According to Treis (2008, pp. 418, 220), Kambaata has a linker **n** before the possessive suffix **-s**, and the emphatic function of **n** is also discussed under the pragmatically determined **-n** morpheme.

[168] Anbessa Teferra's (2014, p. 35) use of the term postposition for **-nni** is vague in such a context. Is it an instrumental or ablative marker in such a construction?

4.1.4.2 Possessive / Genitive Pronouns

Personal possession is expressed by several constructions across the HEC languages. In Burji and Gedeo, third-person possessive pronouns show the same gender-conditioned endings that recur in nominal genitive constructions (§4.1.4.1), suggesting these forms are morphologically complex. Hadiyya similarly uses independent third-person possessive pronouns preceding the possessed noun. In Kambaata and Sidaama, by contrast, third-person possession is marked both by a pronoun and by a suffix on the possessed noun itself (**-si** for masculine, **-se** for feminine possessors).

The list below presents the third-person possessive/genitive forms with illustrative phrases showing their syntactic distribution.[169]

	Genitive	**Example**	**Gloss**
Burji	**ʔisákkụ** / **ʔisákkạ** (3sg.m.nom/acc)	*ʔisákkạ ʔaabboó*	‘his father’
	ʔiʃeékkụ / **ʔiʃeékkạ** (3sg.m.nom/acc)	*ʔiʃeékkạ ʔanaá*	‘her husband’
	ʔisáccị / **ʔisáttạ** (3sg.f.nom/acc)	*ʔisáttạ ʔámạ*	‘his wife’
	ʔiʃeéccị / **ʔiʃeéttạ** (3sg.f.nom/acc)	*ʔiʃeéttạ ʔayyé-ccị*	‘her mother’
Ged	**isike** / **isika** (3sg.m.pred/acc)	*isika belto*	‘his son’
	iseke / **iseka** (3sg.m.pred/acc)	*iseka belto*	‘her son’
	isit’t’e / **isit’t’a** (3sg.f.pred/acc)	*isit’t’e beltoni*	‘his daughter’
	iset’t’e / **iset’t’a** (3sg.f.pred/acc)	*belto iset’t’a*	‘her (own) daughter’
Had	**ʔít’t’i** (3sg.m)	*ʔít’t’i lókko*	‘his leg’
	ʔísi (3sg.f)	*ʔísi lókko*	‘her leg’
	ʔíssi (3sg.hononrific)	*ʔíssi lókko*	‘his/her leg’ (honorific)
Sid	**isí** (3sg.m)	*isí ama*	‘his mother’
	isé (3sg.f)	*isé beettó*	‘her child’
Kam	**isí** (3sg.m.gen)	*ísi-ta béetu*	‘it is his daughter’
	isú (3sg.m.acc)	*isí beetú*	‘his son’
	isé (3sg.f.gen)	*isé beetú*	‘her son’
	iséta (3sg.f.acc)	*isé beetú-ta*	‘her daughter’

[169] Data sources: Hadiyya (Tadesse Sibamo, 2015, p. 88), Burji (Tesfaye Baye, 2015), Gedeo (Eyob Kelemework, 2015. pp. 269, 115), Sidaama (Kawachi, 2007), and Kambaata (Treis, 2008, p.c.).

Gender agreement in possessive constructions varies by language: Some mark possessor gender (Sidaama and Kambaata), and others mark possessed noun gender (Burji and Gedeo) (see §4.1.4.3 for a detailed analysis). In Gedeo, these possessive forms combine with person- and case-conditioned enclitics. Eyob Kelemework (2015, p. 270) identifies three sets: **=ki, =(t')t'i** (nominative), **=ka, =(t')t'a** (accusative), and **=ke, =(t')t'e** (predicative), yielding the surface variants shown in the examples above. Kambaata and Sidaama instead mark possession on the possessed noun with the suffixes **-sí** (3sg.m) and **-sé** (3sg.f).

The first-person possessive form may reflect a compound of the genitive marker **ʔ** and a first-person pronominal element **-y**. This same **-y** element appears in other first-person pronouns (§4.1.6.1), supporting its analysis as a core person morpheme. Forms preceded by a hyphen are suffixal variants:

	[135]		[188]	
PHEC	***ʔ-y**	'my'	***ni-ʔ-**	'our'
Burji	**ʔíyyụ / ʔiíyyạ**	'my'	**ninka/ninku, ninci/ninta**	'our'
PNHEC	***ʔy**	'my'	***niʔ-**	'our'
Had	**ʔi, -i**	'my'	**ni**	'our'
Kam	**íi, -ʔe**	'my'	**nii, -nne**	'our'
Sid	**-ʔya, ʔane**	'my'	**ninké, -nke**	'our'
Ged	**anki / ant'i**	'my'	**noʔoki, noʔot't'i**	'our'

These reconstructions remain provisional due to substantial diversity across daughter languages. Second-person possessive pronouns show similar patterns, with the fused (§3.3.1) and metathesized (§3.3.2) genitive glottal stop as well.

	[253]	2sg.poss	[187]	2pl.poss
PHEC	***ki-ʔ-**	'your' (sg)	***kin-ʔ**	'your' (pl)
Burji	**ʃiyyú/ʃíyyạ, ʃíccị/ʃíttạ**	'your' (sg)	**ʃínkụ/ʃínkạ, ʃínci/ʃíntạ**	'your' (pl)
PNHEC	***ki-ʔ, *-ʔ-ki**	'your' (sg)	***kin-ʔ-** → ***kiʔn-**	'your' (pl)
Had	**ki-**	'your' (sg)	**kiʔnúwwi**	'your' (pl)
Kam	**kíi, -kki**	'your' (sg)	**kiʔneʔee, kiʔnee, -ʔne**	'your' (pl)
Sid	**-kki**	'your' (sg)	**kiʔné**	'your' (pl)
Ged	**ati-ki, ati-t't'i**	'your' (sg)	**(haʔnnoke, haʔnnot't'e)**	'your' (pl)

The Gedeo parallel stem *haʔn-* is generalized from the second-person plural nominative stem [104] *haʔno* 'you' (pl) (see §4.1.6.1).

The third-person pronouns also show the traces of the genitive glottal stop in their interaction with the consonant of the clitic via the regular processes of fusion in Hadiyya and Gedeo and gemination in Burji (§3.3.1).

	[279]	Poss	[366]	Poss	[318]	Poss
PHEC	***isa-ʔ-**	'his'	***ise-ʔ-**	'her'	***isn-ʔ-**	'their'
Burji	**ʔisákkụ/ ʔisákkạ, ʔisáccị/ ʔisáttạ**	'his'	**ʔiʃeékkụ/ ʔiʃeékkạ, ʔiʃeéccị/ ʔiʃeéttạ**	'her'	**ʔisinákkụ/ ʔisinákkạ, ʔisináccị/ ʔisináttạ**	'their'
PNHEC	***isi-ʔ-**	'his'	***ise-ʔ-**	'her'	***isn-ʔ-**	'their'
Had	**issi, ʔíṭ'ṭ'i**	'his/her'	**issi, ʔísi**	'her/his, her'	**ʔissúwwi**	'their'
Kam	**isí**	'his'	**isé**	'her'	**issoʔóo**	'their'
Sid	**-si**	'his'	**-se**	'her'	**-nsa**	'their'
Ged	**isike/isit'ṭ'e**	'his'	**iseke/iset 'ṭ'e**	'her'	**insaʔnnek e/insaʔnn et'ṭ'e**	'their'

The geminated **ss** in the plural Hadiyya cognate *ʔissúwwi* 'their' is from the full assimilation of the stem-final **s** and the plurative **-n** as discussed in §3.1.1.2. This plural stem was subsequently generalized to the singular stems in the masculine and the feminine, but is still transcribed (or alternates) with the single **s** in some sources.

4.1.4.3 Genitive Components

The previous sections discussed the genitive markers. The following paragraphs briefly present word order and gender assignment in genitive constructions.

Word Order

The typical order of the words in genitive constructions in HEC is that the possessor precedes the possessed (Modifier-Head). The possessor receives the genitive marker/clitic. This Modifier-Head (Possessor-Possessed) structure contrasts with the Head-Modifier order common in related Lowland East Cushitic languages. Tosco (1994) argues strongly that this Modifier-Head order in HEC is not original but represents a significant historical shift from an earlier Head-Modifier stage, likely in archaic Proto-East Cushitic. He proposes that this typological restructuring was driven by the grammaticalization of specific constructions, such as the cleft-derived genitives discussed earlier,

which harmonized HEC's nominal syntax with its consistent Verb-Final sentence structure.

In Hadiyya, Hudson's (1976, p. 254) "apposition" and Tadesse Sibamo's (2015, p. 60) "juxtaposition" are used to describe the order within the genitive components and assign it as the only way to express genitive relationships regardless of gender (cf. the subsection on Genitive ʔ). However, since the Modifier-Head is the default genitive component order in addition to the markers in the rest of HEC, word order terms such as apposition and juxtaposition do not tell us much comparatively.

gareec maara	'sheep meat'
ii-ann mentic	'my father's wife'
jallor ooso	'Jalloro's children'
laccaam-i k'armánco	'Lachaame's relative'
ʔantabaʔ-i ládo	'hen's feather'
meent-i horoóre	'hair of women'

Tadesse Sibamo describes the vowel **-i** as a linker in the last three examples above (transcribed as one word in the source).[170] Is it possibly a genitive marker or an epenthetic **-i-**, as it is also attested in the genitive constructions with the glottal stop (see the subsection on Genitive ʔ)?

In Gedeo, according to Eyob Kelemework (2015, pp. 335, 236), the order of the elements in the noun phrase is that "the head noun usually comes on the right-most side, and all the modifiers on its left. Nonetheless, this order of elements is flexible. Especially, if the modifiers are possessive pronouns, demonstratives, and numerals, the head noun may appear before its modifiers". For instance, the attributive nouns or adjectives may also follow the nouns they modify.

belt-i	manj-i-cco-ti=ke	me??-ø-e
child-nom.m	person-ep-sg-loc-gen.m	go-3sg.m-sg.perf

'The son of the man / woman went.'

[170] Tadesse Sibamo (2015) transcribes them as one word. However, they are most likely phrases (compare *maaʔlimíne* 'butchery' with Ritter's (2007) *laʔl mine* 'cattle area (in house)'. Similarly, the first three examples from Hudson).

manj-i-cco-ti=t'i	belto	me??-i-t-e
person-ep-sg-loc-gen.f.nom	child	go-ep-3sg.f-sg.pfv

'The daughter of the man / woman went.'

The above examples show that the genitive attributive singular noun *manjicco* 'person' follows the Modified noun *belto* 'child'. Such a flexible order is also found in Kambaata, but for noun phrases within a sentence (Treis, 2008, p. 102).

Treis (p.c; 2008, p. 102) and Kawachi (2007, p. 7, fn. 10) reject Hudson's (1976, pp. 254–255) claims where he posits that Kambaata and Sidaama ordinarily use apposition only and no morphemic marker to indicate the genitive. In Sidaama, Kawachi regards Hudson's example with the proper nouns *besʃa ooso* 'Besha's children', at least in the dialect of Sidaama he examined, as ungrammatical. It should instead be *besʃ-í ooso* 'Besha's children' with the genitive suffix **-i** of the masculine proper (and common) nouns ending in **-a**.

Gender

Except for in Hadiyya, gender distinction seems central to the genitive construction in HEC. In the Sidaama and Kambaata languages, gender refers to the gender of the possessor. This is in contrast to the genitive with the exponents **k** and **t** in Burji and Gedeo, where the choice of **k** (m) or **t** (f) depends on the gender of the possessed (see Hudson 1976, p. 254; Tesfaye Baye, 2015, p. 117; Eyob Kelemework, 2015, p. 115).

This split in gender assignment (possessor-controlled in Sidaama/Kambaata vs. possessed-controlled in Burji/Gedeo) is notable. If PHEC had a single consistent pattern, the current distribution would require at least one branch to have innovated a reversal. Alternatively, the genitive constructions in the two subgroups may have different historical origins, with Burji/Gedeo retaining a possessed-agreement pattern (comparable to Lowland East Cushitic) while PNHEC innovated a possessor-controlled vowel/suprafix system. This question requires further investigation.

In summary, several processes appear to have shaped the development of genitive markers of the HEC languages. Lenition is evident in the proposed shift from **k** to **h** and potentially to zero in Kambaata and Sidaama (§3.2.2). Assimilation is noted in Burji voicing assimilation (§3.1.1) and potentially in Kambaata vowel harmony within compounds. Coalescence/glottal stop elision

(§3.3.3) is suggested to explain Sidaama and Kambaata vowel/suprafix genitives, in which the stem and the vowels following the lost genitive consonant might have been influenced (§2.4). Epenthesis is proposed for the **-n-** element preceding consonants in Burji/Gedeo and reasonably for others, potentially arising after degemination (§3.4.3). These processes collectively illustrate the dynamic phonological changes influencing morphological forms in HEC.

4.1.4.4 Historical Development of the Genitive

There are two types of genitive construction. The first type of genitive is found without a following clitic; hence, it has no gender association with the head noun of the genitive construction in Hadiyya, Kambaata, and Sidaama. The second type includes the genitive suffix ***-ʔ**, followed by one of the gender clitics ***=k** (m), and ***=t** (f), which in turn are preceded by epenthetic **-n-** in the common nouns. Kambaata and Hadiyya show evidence of the first type only, while Sidaama has reflexes of both types, and Burji and Gedeo show the second type only.

The genitive ***-ʔ** is retained in Hadiyya in the form of fusion (§3.3.1) and metathesis (§3.3.2) depending on the sonority of the stem-final consonant. It is still evident in the glottalized consonant of the genitive clitic **=t't'** (§3.3.1) in Gedeo. The compensatory vowel doubling and accentuation **-V́(V)** (gen) (§2.2) in Kambaata and Sidaama is further evidence suggesting the loss of the following glottal stop of the genitive (see §2.4 and the subsection on Genitive ʔ).[171] Additionally, the gemination of the following consonant of the clitic **t** (f) and **k** (m) suggests a fusion of the **ʔ** (gen) and the following consonant. It is retained as such (**tt** (gen.f) and **kk** (gen.m)) on proper nouns and pronouns in Burji and Gedeo but was otherwise degeminated into a single consonant (§3.1.1). Consequently, the epenthetic **-n-** occupied the place of the first consonant, i.e., **n=t** (gen=f) and **n=k** (gen=m) upon degemination of **tt** (gen.f) and **kk** (gen.m) on Burji and Gedeo common nouns and Sidaama genitives with clitic **k** (m) and **t** (f) (§3.4.3). This means that all three languages share the fusion/assimilation stage, which yielded the geminates ***-kk** (gen.m) and ***-tt** (gen.f), and which in turn encode the older genitive and the clitic construction ***-ʔ=k** (gen=m) and ***-ʔ=t** (gen=f). Therefore, the current epenthetic **-n-** before the clitic on the common nouns holds the place of the

[171] Treis (2008, p. 37, fn. 23) reported a puzzling instance of a sequence **ʔ**r found on one of the (shared) lexemes where **ʔ** cannot be analyzed as a middle marker: in Kambaata *gorʔá* 'green, unripe berry' and Sidaama *goʔra* 'wild berry similar to blackberry and raspberry' (cited from Kawachi 2007, p. 29, fn. 1).

old genitive ***-ʔ** in the genitive construction (i.e., ***n=k** (gen=m) and ***n=t** (gen=f)). Below is a summary of these processes as attested across the different languages (Assim. = assimilation (1 first, 2 second assimilation), Gem. = gemination, Degem. = degemination, Epen. = epenthesis, Meta. = metathesis).

	Genitive + final/Clitic	Assim.1 + Gem.	Fusion + Gem.	Degem. / Epen.	Assim.2 /Meta.
Bur, Ged, Sid	***ʔ + *k →**	**kk**	**-**	**nk**	**nn**
Bur, Sid	***ʔ + *t →**	**tt**	**-**	**nt**	**nn**
Ged, Had	***t + *ʔ →**	**-**	**t't' (← *t+ʔ)**	**nC'**	**-**
Had	***$C_{i[-son]}$ + *ʔ →**	**-**	**C'C'**	**-**	**-**
Had	***$C_{i[+son]}$ + *ʔ →**	**-**	**-**	**-**	**ʔC**
Sid, Kam	***V_i + *ʔ →**	**-**	**-**	**V́(V)**	**-**

This schema proposes the diachronic pathways but should be read with caution: Not all intermediate stages are directly attested, and the ordering of processes is inferred from synchronic distributions. The core claim is that PHEC ***-ʔ** combined with clitics ***=k** (m) and ***=t** (f) and subsequently underwent assimilatory/fusion processes. It is an arguable interpretation based on the independent evidence from Hadiyya glottalization, Gedeo glottalized **t't'**, and Kambaata/Sidaama compensatory doubling/accentuation. The **-n-** epenthesis is best understood as a repair strategy following degemination, filling the position vacated by the lost first geminate component.

A second assimilation (Assim.2) ***nk** / ***nt** → **nn** is unique to Sidaama.[172] Additionally, Sidaama has a following **-i-**, which is inserted in **-nn-i-** to avoid the **CCC** cluster that would result when a following consonant-initial morpheme (e.g., clitic) is attached.

Almost all the retained forms of the genitive construction show evidence of the glottal stop. Treis (2008, p. 226; fn. 134) notes that in Kambaata, the linker **n** might also be called a grammatically determined epenthetic consonant. Although that is not applicable to the Kambaata genitive, historically, it aligns with the morphophonological epenthesis process of **-n-** (§3.4.3) after the (***-kk-**, ***-tt-**, ***-t't'-**) degemination and may apply to most, if not all, the instances of **-n-** presented above on the common nouns in the genitive

[172] The assimilation patterns 1 and 2 are not improbable (§3.1.1); for instance, Gedeo **ʔt** assimilating to **tt** in *mado?-* 'wound' → *madot-t-e* (wound-3sg.f.perf) (Eyob Kelemework, 2015, p. 62). However, it is unclear whether all languages have synchronic morphophonemic evidence or whether we must rely solely on diachronic explanations. The same applies for Sidaama ***nk** / ***nt** → **nn**.

construction. The elements involved in that process are the genitive ***-ʔ**, stem-final $\mathbf{C_i}$, $\mathbf{V_i}$, and the clitic ***k**/***t** (Bur = Burji).

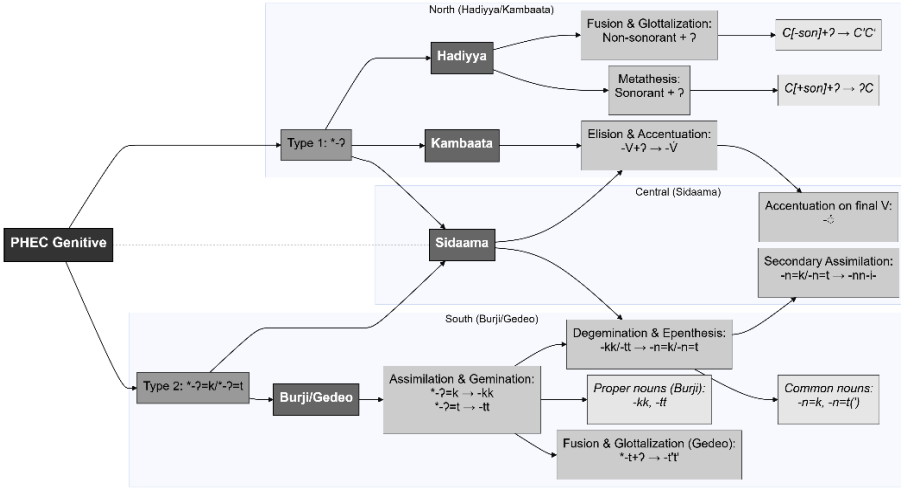


Figure 3: *Historical Development of the Genitive*

The comparative data reveal a typological and geographical cline in HEC genitive constructions: from the morphologically complex systems of Burji and Gedeo in the south through the central mixed system of Sidaama, to the simple system of Hadiyya and Kambaata in the north. Burji and Gedeo present the most complex systems in a robust paradigm that distinguishes case (nominative/accusative/predicative), gender (masculine/feminine), and noun type (common/proper). The variation in these exponents' retention and function reflects broader language change processes in the HEC languages, including morphological restructuring, functional reanalysis, and phonological attrition.[173]

The above reconstructions are considered arguable proposals because the diversity in the genitives might alternatively suggest independent innovations in each language. However, such an alternative interpretation ignores the comparable parallels in Burji and Gedeo, on the one hand, and Kambaata and

[173] This reconstruction, particularly the proposed origin of the **k**/**t** clitics from copulas in cleft constructions leading to a word order shift, aligns with the analysis by Tosco (1994). It differs from Hetzron's (1980) hypothesis, which, while acknowledging the **k**/**t** agreement as an important link between HEC and Agaw, suggested these markers were transferred from the possessed noun to the possessor. Tosco argues that the cleft-origin model better accounts for the attested structures without invoking such a transfer.

Sidaama, on the other hand. Therefore, the proposed reconstructions could be solidified upon further research into individual languages (especially Hadiyya), more harmonized analysis, and use of the terminology where languages exhibit comparable systems.

4.1.5 Gender and Case

HEC nominal morphology treats gender and case as formally intertwined categories. The masculine marker ***-k** and feminine marker ***-t** recur across demonstratives, definiteness markers, and pronominal endings; the case vowels ***-i** (nominative), ***-o** (masculine accusative), and ***-e** (feminine accusative) attach directly to these gender exponents or appear fossilized in pronominal paradigms. This section surveys the distribution of both markers and evaluates what can be reconstructed to PHEC or even Pre-PHEC, beginning with gender (§4.1.5.1) and then turning to case (§4.1.5.2).

4.1.5.1 Gender

The masculine marker ***-k** and feminine marker ***-t** are pan-Cushitic exponents that recur across the HEC languages in demonstratives, definiteness markers, case suffixes, and agreement morphology. Their distribution supports reconstruction to PHEC with secure confidence; the individual languages differ only in which grammatical functions these consonants continue to serve productively.[174]

[174] Data sources: Burji (Tesfaye Baye, 2015, pp. 100, 104), Hadiyya (Hudson, 1976, p. 251; Tadesse Sibamo, 2015, p. 53), Kambaata (Treis, 2008, pp. 108, 129–130, 361), Sidaama (Kawachi, 2007, pp. 200, 483), Gedeo (Dawit Tilahun, 2019, pp. 91, 112; Gasparini, 1995, pp. 9, 2).

	-k (m)	-t (f)	Function
Burji	**-ʃi̱**	**-tteé**	gender
	kú	**-cí**	demonstrative ‘this’
	-gu	**-ʃi**	subject definite
Had	***-k**	**-ta**	gender
	kúki	**túte**	definiteness demonstrative ‘this’
Kam	**-ha, -hu**	**-ta, -ti**	secondary case / gender
	ka, ku	**ta, ti**	demonstrative attribute ‘this’
Sid	**ko**	**te**	adnominal demonstrative ‘this’
	-ho	**-te**	dative case (Unmodified common nouns)
Ged	**koo**	**tee**	vocative (proper nouns)
	kunni	**tinni**	demonstrative pronoun ‘this’
	-ki	**-ti**	definite (article)
	-k-eʔe	**-t-eʔe**	dative case (common nouns)

Hadiyya **-ta** is found on some nouns (Hudson, 1976, p. 251), while ***-k** is only recoverable from underlying forms via a morphophonological process triggered by the attachment of some suffixes (§3.3.4). In Kambaata, the vowels following the gender exponents are case markers: **-a** for the accusative and oblique and **-i** (f) or **-u** (m) for the nominative (Treis, 2008, pp. 127, 322). Treis (2008, p. 108) analyzes the development of the above functions in Kambaata as a “recent grammaticalization of case- and gender-sensitive demonstratives into secondary case / gender markers”. Yri & Pepper (2019) specify that, in Sidaama, the vowels **-u** and **-o** follow the masculine gender consonant **k-** and the vowels **-i** and **-e** follow the corresponding consonant **t** of the feminine in the demonstrative pronouns, but they do not associate them with a specific case as in the other HEC languages. The Gedeo definite markers **-ki** and **-ti** become **-ka**, **-ta** in the accusative and **-ke**, **-te** in the other cases (Gasparini, 1995, p. 2).

Hudson (1976, p. 251) lists some of the occasions when the gender exponents occur on the HEC nouns: (a) in their agreement with verbs, (b) in their choice of demonstratives, (c) in their construction of the genitive (in some languages, §4.1.4.1), and (d) in their choice of a suffix (article), to some extent.

To summarize: PHEC ***-k** (masculine) and ***-t** (feminine) are securely reconstructed on the basis of uniform attestation across demonstratives, definiteness markers, and dative/genitive case morphology. The consonants themselves are stable; the vowels that follow them encode case and other grammatical distinctions, which are addressed in §4.1.5.2. The pan-Cushitic

attestation of ***-k** / ***-t** (Tosco, 2020, p. 294; Hudson, 1976, p. 246) situates the HEC pattern within the broader Afroasiatic typology of gender marking.

4.1.5.2 Case

The terms nominative and accusative (some sources refer to it as the absolutive) refer primarily to form; subject and object denote function. This distinction matters because the pronominal paradigms preserve case marking that is no longer productive in standard nominal inflection (§1.5.2).

As an alternative to the null hypothesis that the current terminal vowels have nothing to do with any fossilized case marking, one could argue for potential old case functions that the TVs used to serve, based on remnants of synchronic evidence (see §2.2.3).

Accusative

The accusative case markers, reconstructed as ***-o** (← ***-ó**) for masculine and ***-e** (← ***-é**) for feminine, survive primarily in fossilized pronominal forms rather than in productive nominal paradigms. Their distribution is most clearly visible in the third-person singular object pronouns and in the interrogative. The following subsection presents the evidence for these reconstructions and notes where individual languages have generalized or restructured the inherited pattern.

Accusative *-e and *-o

The best evidence for the accusative (**-e** (f) and **-o** (m)) vowels comes from the singular personal pronouns. The fossilized feminine accusative ***-e** is reconstructed in the pronouns [151] **is-e* ‘her’ and [28] **ay-e* ‘who?’. The nominative counterparts are [6] **is-i* ‘she’ and [28] **ay-i* ‘who?’ (→ Kam and Sid *ayi* (nom) ‘who?’).

	[151]		**[28]**		**[292]**	
PHEC	***ise**	‘her’	***aye**	‘who?’	***iso**	‘him’
Burji	**ʔiʃeé**	‘her’	**ayyee**	‘who?’	**ʔisi̯**	‘him’
Had	**ʔíse̯**	‘her’	**ayye**	‘who?’	**it’t’o**	‘him’
Kam	**isé-ta**	‘her’	**aye**	‘who?’	**isú**	‘him’
Sid	**ʔisé**	‘her’	**aye**	‘who?’	**ʔisó**	‘him’
Ged	**isee, iseeʔe**	‘her’	**ayye**	‘who?’	**isoo, isooʔo**	‘him’

The masculine counterpart **-o** of the feminine accusative **-e** above is also inferred as an arguably fossilized suffix from the pronoun [292] PHEC **is-o* 'him'. Kambaata *isú* 'him' shows a regular **-u**. In Burji, the vowel **-o** of the object pronoun has been restructured due to analogy with the vowel of the subject pronouns [7] *ʔísi̯* 'he, it' and [6] *ʔíʃi̯* 'she', thus becoming *ʔisi̯* 'him'.

In set [407], **ebel-, *ebal-* 'so-and-so' behaves similarly to pronouns with the TVs **-o** and **-e** in Hadiyya and Kambaata. In other words, they show gender distinction between the masculine accusative Hadiyya *ebaro*, Kambaata *ebalo* and the feminine accusative Hadiyya *ebare* and Kambaata *ebale* 'so-and-so', even though Kambaata **-o** is an irregular reflex as noted in the terminal vowel **-o** section in §2.2.3.1.

In addition to the pronouns, the demonstratives (pronouns/determiners/adjectives) provide fossilized evidence of the arguably proposed accusative ***-o** and ***-e**.

		Masculine Acc	**Feminine Acc**	**Gloss**
Burji	[418]	**ka**	**ta**	'this'
Had		**ka**	**ta**	'this' (det)
Kam		**ka**	**ta**	'this' (det)
Sid		**ko, konne, koe**	**te, tenne**	'this'
Ged		**konne**	**tenne**	'this' (det)
Burji	[281]	**ci**	**ci**	'these'
Had		**kakeen**	**kakeen**	'these'
Kam		**kara**	**-**	'these' (pron)
Sid		**konne, ko(ʔo)**	**tenne, te(ʔe)**	'these'
Ged		**konneera**	**tenneera**	'these'

The TV is mainly **-a** in Burji, Hadiyya, and Kambaata. However, Sidaama and Gedeo may have maintained a distinction in vowel quality with **-o** for the masculine and **-e** for the feminine before the attachment of **-nnV(V)**.

The singular demonstrative [418] can be reconstructed as the masculine accusative ***k-o** and feminine accusative ***t-e** (f.acc) 'this'. This reconstruction assumes that the Gedeo and Sidaama forms have preserved a more archaic vowel contrast **-o** (m) and **-e** (f). However, such an assumption disregards the fact that the gender distinction is already realized by the consonants ***k** (m) and ***t** (f), making the gender distinction that the vowels are supposed to realize in addition to the case marking redundant. Therefore, an alternative reconstruction of the accusative vowel would be accusative case ***-a** based

on the Burji, Hadiyya, and Kambaata reflexes, with the gender realized via the consonant contrast, i.e., ***ka** (m.acc) and ***ta** (f.acc).[175] One might argue in favor of the first reconstruction that the consonantal gender distinction of the demonstratives might have encouraged the later neutralization of the vowel distinction, or that the vowel **-a** is a result of vowel coalescence: **ko** and **te** + **a** → **ka** and **ta** in Burji, Hadiyya, and Kambaata. The pronouns provide further evidence, in support of the first reconstruction, for maintaining ***-e** (f) and ***-o** (m) in the accusative.

The following examples are identical to the ones presented under the section on variation in terminal vowels (2.2.3.2). However, these examples include the respective gender of the stem in each language to examine if the TV of the nouns matches the proposed genders (gen) in the accusative, assuming that some might be fossilized case markers:

	PHEC	Gloss	Had	Sid	Ged	Kam	Burji
[174]	*hamas-	'snake'	**-aʃʃa**	**-o**	**-me**	**-u**	**-i**
			-	(m)	(f)	(m)	(m)
[137]	*lal-	'cattle'	**-o**	**-o**	**-o, -i**	**-u**	**-i**
			-	(m)	(m)	(m)	(m)
[444]	*bagaz-	'spear'	**-o**	**-o**	**-o**	**-u**	**-i**
			-	(m)	(m)	(m)	(m)
[95]	*ʃomb-	'lung'	**-**	**-o**	**-o**	**-**	**-i**
			-	(m)	(m)	-	(m)
[606]	*gubeed-	'thigh'	**-o**	**-a**	**-a**	**-**	**-i**
			-	(m)	-	-	-
[9]	*arrab-	'tongue'	**-o**	**-o**	**-o**	**-i-ta**	**-a**
			-	(m)	(m)	(f)	(f/m)
[4]	*gulib-, *gilub-	'knee'	***-co**	**-co, -e**	**-o**	**-i-ta**	**-a**
			-	(m)	(m)	(f)	(f)
[382]	*oll-	'neighbour'	**-aʔa**	**-a**	**-aʔa**	**-oo**	**-a**
			(pl)	(m)	(m)	(m)	(m/f)
[51]	*lokk-, *lekk-	'foot, leg'	**-o**	**-a**	**-a**	**-a-ta**	**-a**
			-	(f)	(f)	(f)	(f)
[293]	*waʔ-	'water'	**-o**	**-a**	**-aʔe, -eʔe**	**-a**	**-a**
			-	(m)	(m)	(m)	(m)
[88]	*gudum-	'shoulder'	**-o**	**-**	**-o**	**-u**	**-a**
			-	-	(f)	(m)	(f/m)
[195]	*fal-, *falfal-	'omen, magic'	**-o**	**-**	**-o**	**-a**	**-aa**
			-	-	(f)	(m)	(m)
[5]	*san-	'nose'	**-e**	**-o**	**-o**	**-u-ta**	**-a**

[175] Blažek & Malášková (2016) identify the final vowel **-a** on the first-person object pronouns as the absolute case suffix: Oromo Tulema *ánā*, *na*, Konso *ána*, and Mashile *ana*. As mentioned above, "absolutive" is used interchangeably with "accusative" in the HEC literature.

	PHEC	Gloss	Had	Sid	Ged	Kam	Burji
			-	(f)	(f)	(f)	(f)
[439]	*reħ-	'death'	**-o**	**-o**	**-o**	**-u-t**	**-a**
			-	(f)	(f)	(f)	(f)
[206]	*awd-	'threshing floor'	**-a**	**-umme**	**-e**	**-u**	**-a**
			-	(f)	(f)	(m)	(f)
[408]	*mut-	'awl for basketwork'	**-a**	**-C'C'e**	**-a**	**-a**	**-a**
			-	(f)	-	(m)	-
[381]	*k'ul-, *k'ul~k'ul~l-ʔ-	'leech'	**-a**	**-e**	**-a**	**-a**	**-ee**
			(pl)	(f)	-	-	(m)
[1]	*ill-	'eye'	**-e**	**-e**	**-e**	**-i-ta**	**-a**
			-	(f)	(f)	(f)	(f)
[415]	*afal-	'liver'	**-e**	**-e**	**-**	**-i-ta**	**-a**
			-	(f)	-	(f)	(f)
[54]	*min-	'house'	**-e**	**-e**	**-e**	**-i**	**-a**
			-	(m)	(m)	(m)	(m)
[550]	*ibb-	'louse'	**-a**	**-e, -iicco**	**-e**	**-i-ta**	**-a**
			-	(f)	(f)	(f)	(f)
[224]	*t'ur-	'dirt of body, clothing'	**-a**	**-e**	**-e**	**-i**	**-ee**
			-	(m)	(m)	(m)	(f)
[544]	*c'ub- (PNHEC)	'dagger'	**-eʔe**	**-e, -a**	**-e**	**-i-ta**	
			-	(f)	-	(f)	

The above examples show the respective gender (masculine (m), feminine (f)) underneath the terminal vowel of each noun. Those with only a dash "**-**" indicate absence of gender identification in the source (e.g., all Hadiyya examples) or lack of a cognate. The available Hadiyya sources do not indicate the gender of the nouns. Kambaata **-ta** is the feminine accusative suffix, and it is likely added after the TV (mostly **-i** and **-u**). Some examples include additional suffixes, such as the singulative or derivational suffix in [174], [550], and [4]. In [408], Sidaama **-C'C'** includes a glottalized stem-final **-t't'-** probably triggered by a glottal suffix (§3.3.1). In the examples where there is more than one TV for the same stem in the same language, e.g., Sidaama **-e, -a** in [544], and Gedeo **-o, -i** in [137], each is from a different data source.

The following paragraphs address the question of whether there is any association between the quality of the terminal vowel and the assigned gender (i.e., a fossilized marker) or whether the gender is inherent to the root.

Concerning the first half of the question, the data show that a single stem can end in **-a**, **-i**, **-o** (Kambaata **-u**), or **-e**, depending on the language, e.g., [606]

Hadiyya **-o**, Gedeo and Sidaama **-a**, Burji **-i**. In general, the observed NHEC patterns in terminal vowel quality and gender match, including masculine (m) forms that show a back TV (**-o**), while feminine forms surface with a front TV (**-e**). The corresponding vowels in Kambaata are **-u** and **-i**, respectively. In contrast to Kambaata, Burji shows **-i** for the masculine. The TV **-a** appears on both masculine and feminine stems in all HEC languages. The following examples show regular correspondences aligning with these patterns for the masculine and feminine found in 8 out of the total 23 examples:

	Had		**Ged**		**Sid**		**Kam**		**Burji**	
[95]	**-**	-	**-o**	(m)	**-o**	(m)	**-**	-	**-i**	(m)
[174]	**-**	-	**-**	-	**-o**	(m)	(m)	(m)	**-i**	(m)
[444]	**-o**	-	**-o**	(m)	**-o**	(m)	**-u**	(m)	**-i**	(m)
[137]	**-o**	-	**-o, (-i)**	(m)	**-o**	(m)	**-u**	(m)	**-i**	(m)
[415]	**-e**	-	**-**	-	**-e**	(f)	**-i-ta**	(f)	**-a**	(f)
[1]	**-e**	-	**-e**	(f)	**-e**	(f)	**-i-ta**	(f)	**-a**	(f)
[550]	**(-a)**	-	**-e**	f	**-e**	(f)	**-i-ta**	(f)	**-a**	(f)
[544]	**-eʔe**	-	**-e**	f	**-e, (-a)**	(f)	**-i-ta**	(f)	**-**	-

Overall, Kambaata exhibits the highest degree of gender and vowel quality matching with **-i** in the feminine and **-u** in the masculine stems in the above and other examples.

However, there is counterevidence for the above patterns, in which back vowels occur in feminine stems and front vowels occur in masculine stems.

	Had		**Ged**		**Sid**		**Kam**		**Burji**	
[224]	**-a**	-	**-e**	(m)	**-e**	m	**-i**	(m)	**-ee**	(f)
[54]	**-e**	-	**-e**	(m)	**-e**	m	**-i**	(m)	**-a**	(m)
[439]	**-o**	-	**-o**	(f)	**-o**	(f)	**-u-t**	(f)	**-a**	(f)
[5]	**-e**	-	**-o**	(f)	**-o**	(f)	**-u-ta**	(f)	**-a**	(f)

As mentioned above, the low TV **-a** is found in both masculine and feminine stems (see §4.1.5.1).

	Had		Ged		Sid		Kam		Burji	
[51]	**(-o)**	-	**-a**	f	**-a**	f	**-a-ta**	f	**-a**	f
[293]	**(-o)**	-	**-aʔe, -eʔe**	m	**-a**	m	**-a**	m	**-a**	m
[382]	**-aʔa**	pl	**-aʔa**	m	**-a**	m	**(-oo)**	m	**-a**	m/f
[408]	**-a**	-	**-a**	-	**-**	-	**-a**	m	**-a**	-
[381]	**-a**	pl	**-a**	-	**(-e)**	f	**-a**	-	**(-ee)**	m

The TVs that do not fit the quality of the ascribed pattern are placed between parentheses, e.g., **-e** in [381], **-o** in [51], and **-a** in the previous examples [550] and [544].

Overall, the terminal vowels of the PHEC accusative pronouns **is-e* (3sg.f.acc) 'her' and **is-o* (3sg.m.acc) 'him' coincide with the high frequency of terminal **-o** and **-e** in the absolutive/accusative nouns in at least Gedeo, Hadiyya, and Sidaama. They are also the only feature of the PNHEC noun root **ar-* that distinguishes gender and sex between [212] **ar-o* (m) 'husband' and [271] **ar-e* (f) 'wife'.

Accusative *-é and *-ó

Accented ***-é** (f.acc) and ***-ó** (m.acc) could be reconstructed for the PHEC accusative case as the synchronic evidence proposes. Prior to the current accented single vowels, the accusative might have been double ***-ée** (f.acc) and ***-óo** (m.acc) or vowel and consonant sequences of the shape ***-V́C**, representing a valid accent-bearing syllable, as discussed in §2.2.

The alternative scenario (null hypothesis) is that the gender is inherited from the root. Such a scenario requires regularity in the assigned gender of each cognate set, regardless of the quality of the TV in the individual languages. Half of the 23 examples listed above have the same gender across all languages in which they are attested, regardless of TV.

	PHEC	Gloss	Gender
[206]	***awd-**	‘threshing floor’	(f)
[550]	***ibb-**	‘louse’	(f)
[415]	***afal-**	‘liver’	(f)
[1]	***ill-**	‘eye’	(f)
[4]	***gulib-**	‘knee’	(f)
[51]	***lokk-, *lekk-**	‘foot, leg’	(f)
[439]	***reħ-**	‘death’	(f)
[54]	***min-**	‘house’	(m)
[293]	***waʔ-**	‘water’	(m)
[95]	***ʃomb-**	‘lung’	(m)
[137]	***lal-**	‘cattle’	(m)
[444]	***bagaz-**	‘spear’	(m)
[382]	***oll-**	‘neighbour’	(m)
[224]	***t’ur-**	‘dirt of body, clothing’	(m)

There are two exceptions in the Burji examples, where [382] is assigned masculine in one source and feminine in another, and where [224] is feminine in contrast to its cognates in the rest of HEC.

In summary, the immediate accusative reconstruction posits ***-e** (f.acc) and ***-o** (m.acc) based on the surface forms of third-person pronouns and the terminal vowel patterns in nouns. The more abstract layer (accented ***-é** and ***-ó**) accounts for phontactic properties (accent position, syllable weight) that the unaccented single vowels alone do not explain. Therefore, they are traceable to earlier double vowels or vowel-consonant sequences pending broader East Cushitic confirmation. These competing hypotheses do not affect the core reconstruction, which is arguable at the PHEC level.

Nominative

The nominative case markers, reconstructed as ***-i** for feminine and ***-u** for masculine, are preserved in third-person subject pronouns and demonstratives. Unlike the accusative, where the vowel contrast is clearly attested across most HEC languages, the nominative markers show considerable leveling: many languages generalize ***-i** to both genders, obscuring the original distinction. The evidence for ***-u** as the masculine nominative is strongest in Burji, Kambaata, and Sidaama demonstratives;

elsewhere, it has been replaced by analogy. The following subsections present the immediate evidence for ***-i** and ***-u**, then propose a more abstract reconstruction involving glides ***-y** and ***-w** to account for phonological properties such as devoicing and lack of accent.

Nominative *-i and *-u

Terminal **-i** is most frequently found on nominative pronominals in all HEC languages: Gedeo (Gasparini, 1995, p. 1), Hadiyya (Tadesse Sibamo, 2015, p. 59), Burji (Tesfaye Baye, 2015, pp. 48, 104, 115), Sidaama (Anbessa Teferra, 2014, p. 31; Indrias Xa'miso et al., 2007, p. 637), Kambaata (Treis, 2008, pp. 322, 130, 127). Thus, nominative ***-i** is arguably reconstructable to PHEC. It is most likely the feminine nominative marker in the third-person subject pronouns, e.g., PHEC [6] **isi* 'she', equivalent to the feminine accusative ***-e** in PHEC [151] **ise* 'her'.

The back counterpart **-u** is found as the terminal vowel of nominative forms in various nominals in Burji (Tesfaye Baye, 2015, pp. 48, 104, 115), Kambaata (Treis, 2008, pp. 322, 130, 127), and Sidaama (Anbessa Teferra, 2014, p. 31; Indrias Xa'miso et al., 2007, p. 637) but never attested in Gedeo or Hadiyya. It alternates with **-i** in Sidaama. In Burji, Kambaata, and Sidaama, there is a tendency to associate **-u** with the masculine nominative in terms of gender. Thus, it is arguably reconstructed as a remnant marker of (pre-)PHEC ***-u** (m.nom) equivalent to the feminine ***-i** (f.nom). However, the latter has likely been generalized by analogy, resulting in frequent attestation of **-i** in both the masculine and feminine nominative.

The demonstratives (pronouns/determiners/adjectives) provide the best fossilized evidence of the tentatively proposed nominative masculine ***-u** and arguable feminine ***-i**, among the pronominals.[176]

[176] Sources of additional examples (singular and plural): Sidaama (Kawachi, 2007, p. 202), Burji (Hudson, 1976, p. 256), Gedeo (Dawit Tilahun, 2019, p. 143).

		Masculine Nom	Feminine Nom	
[418]	Burji	**ku**	**ci**	‘this’ (det)
	Had	**ku(k)**	**tu(t)**	‘this’ (det)
	Kam	**ku**	**ti**	‘this’ (det)
	Sid	**ko, kuni, kuyi**	**te, tini, tiyi**	‘this’
	Ged	**kunni**	**tinni**	‘this’ (det)
[281]	Burji	**kan(d)’u-gu**	**cin(d)’i-ʃi**	‘these’ (pron)
	Had	**kukeen**	**tuteen, tu keen**	‘these’
	Kam	**kuru**	**-**	‘these’ (pron)
	Sid	**kuni, kuri**	**tini, kuri**	‘these’
	Ged	**kunniiri**	**tinniiri**	‘these’

The singular demonstrative [418] can be reconstructed as masculine nominative ***k-u** (m.nom) and feminine nominative ***t-i** (f.nom) ‘this’. Such a reconstruction is supported by the reflexes of Burji, Kambaata, and one of the Sidaama variants. The vowel distinction leveling to **-u** in Hadiyya might be due to the redundancy of gender assignment via the consonants ***k** (m) and ***t** (f) and the vowels ***-u** (m) and ***-i** (f) at the same time. The same applies to the Sidaama and Gedeo variants, with **-i** being generalized after the addition of the nasal **-n** (or glide **-y**).

The plural counterparts of the demonstratives in [281] are provided to give more context. They also help explain the nasal **-n** in the Sidaama and Gedeo singular forms, which are a generalization of the plurative forms (see plurative ***-n** in §4.1.1.2).

To summarize the immediate reconstruction: PHEC ***-i** (f.nom) is attested across all HEC languages, appearing on third-person feminine subject pronouns, e.g., [6] **is-i* ‘she’) and generalized to other nominative contexts. PHEC ***-u** (m.nom) is more restricted and preserved in masculine demonstratives, e.g., [418] **k-u* ‘this’ (m) and some masculine nominative pronouns in Burji, Kambaata, and Sidaama, but leveled to ***-i** in most other cells.

Nominative *-y and *-w

The nominative vowels are widely reported as unaccented and devoiced (§2.2.3), even sometimes inaudible in comparison to the accusative. This could be addressed with further abstraction, suggesting that instead of or prior

to the vowels ***-i** (f) and ***-u** (m), there might have been consonant glides ***-y** (f) and ***-w** (m) marking the third-person singular nominative.

	[6]		[7]		[418]	
PHEC	***is-y** → ***is-i**	‘she’	***is-w** → ***is-u**	‘he’	***k-w** → ***k-u**	‘this’ (m)
Burji	**ʔíǰ̣**	‘she’	**ʔísị̣**	‘he, it’	**ku**	‘this’ (det)
Had	**ʔísị̣**	‘she’	**isse, issoʔi, it’t’o**	‘he/she’ (honorific), ‘he’	**ku(k)**	‘this’ (det)
Kam	**íse**	‘she’	**ís, ísu**	‘he’	**ku**	‘this’ (det)
Sid	**ise**	‘she’	**ʔísi**	‘he’	**ko, kuni**	‘this’
Ged	**ise**	‘she’	**isi**	‘he’	**kunni**	‘this’ (det)

In the nominative, the HEC languages generalized the ***-y** from the feminine to the masculine. Kambaata *isu* ‘he’ has a terminal **-u** due to analogy with the accusative counterpart *ísu* ‘him’. However, almost all the reflexes of the masculine nominative demonstrative have retained the expected vowel quality **-u**.

The third-person plural adds ***-n** to the singular stem.

	[11]	
PHEC	***is-n-**	‘they’ (subj)
Burji	**ʔisínụ̣**	‘they’ (subj)
Had	**it’t’uw, issuw**	‘they’ (subj)
Kam	**issoʔóot**	‘they’ (subj)
Sid	**ʔínsa**	‘they’ (subj)
Ged	**insaʔnne, insaʔne**	‘they’ (subj)

The reconstruction of the plural above is tentative due to variation and innovations across the individual languages. There is no gender distinction synchronically in the third-person plural.[177]

The further abstraction proposes pre-PHEC ***-y** (f.nom) and ***-w** (m.nom) as consonant glides that later vocalized to ***-i** and ***-u**, respectively. This layer accounts for several otherwise puzzling features: (a) the characteristic devoicing of nominative vowels across HEC (voiced consonants are less sonorous than vowels, yielding devoiced reflexes upon vocalization); (b) the absence of accent on nominative terminal vowels (the earlier glides were

[177] The glide **-w** found in Haddiyya *it’t’uw*, *issuw* ‘they’ (subj) might be a plurative suffix (§4.1.1.2) rather than the proposed accusative **-w**.

consonants and thus not able to be accented); (c) the palatalized reflexes **ʃ** and **c** in Burji, explainable as the fusion of ***k** + ***y**, ***t** + ***y** (§§3.2.1, 3.1.2.1), and ***s** + ***y**;[178] and (d) parallels with other East Cushitic languages that preserve ***-y** in cognate pronominal forms. The glide hypothesis does not replace the immediate reconstruction but extends it to an earlier stage, providing a principled explanation for the phonological asymmetry between accusative (accented, voiced) and nominative (unaccented, devoiced) markers.

The TV of the nominative, ***-i**, has not only been generalized to both masculine and feminine but also to the other singular subject pronouns [2] **an-i* 'I' and [27] **at-i* 'you'. Therefore, Hudson (2006, p. 296) notes that the HEC languages generally mark nominatives with **-i**, while acknowledging that case marking is complex, is perhaps undergoing change, and needs more research. Beyond HEC, according to Blažek and Malášková (2016), the terminal vowels of the first-person PEC form **ʔani* and the variant **ʔanu*, found in Saho-Afar and Dullay (cf. also Beja *anū*), are noun subject case markers.

In concordance with the above proposed case markers, Hudson (1976, p. 253) reports that in Burji, the "alternation absolute form final e / nominative i, absolute form final o / nominative u is completely regular in proper names […]. Nouns in the accusative case appear in the absolute form, usually final o, a, or e."[179] This suggests a preservation of fully contrasting vowel reflexes of all the reconstructed case exponents in the proper nouns in Burji.

According to Treis (2008, p. 108), "Sasse's observation (1984:111) that the distinction between subject and object is often abandoned in Cushitic languages due to the interference with other grammatical categories (gender and focus) does not hold for Kambaata." The other grammatical categories Sasse mentioned have complicated the synchronic reality, but Treis's statement is accurate at least with regard to HEC, as shown in the above discussion.

[178] For example, in Burji [27] **at-y* becomes *ʔáʃi̯* 'you' (2sg.nom), [418] **t-y* becomes *ci* 'this' (sg.f.det), and [6] **is-y* 'she' becomes **iʃi̯*, i.e., the **-y** triggers the palatalization instead of the front vowel **-i**.
[179] In Konso, only names and pronouns show case (Ongaye Oda, 2013, p. 133).

If one discards all the above proposals of old functions, the terminal vowels on the pronominals have to be interpreted as having nothing to do with case marking (null hypothesis, see §2.2.3).

The case markers can be reconstructed at two levels. At the immediate PHEC level, based on direct surface correspondences, the accusative is marked by ***-e** (feminine) and **-o** (masculine), and the nominative by ***-i** (feminine) and ***-u** (masculine). These reconstructions rest on evidence from third-person pronouns, demonstratives, and terminal vowel patterns in nouns. However, several anomalies complicate this picture: Burji shows paradigm leveling toward ***-i** in both nominative and accusative masculine forms; Kambaata reflects ***-o** as ***-u** (a regular sound change) and ***-e** as ***-i** in some contexts; and the nominative ***-u** is widely replaced by generalized ***-i** outside of demonstratives.

A more abstract analysis can relate the case-vowel asymmetry to earlier phonological material. One possibility is that the accusative vowels continue an older accented opposition (**-é** for feminine vs. **-ó** for masculine), or derive from earlier VV/VC sequences. The nominative vowels are also compatible with an origin in pre-PHEC glides (**-y** for feminine and **-w** for masculine), which would account for their devoicing, lack of accent, and the palatalized consonants found in Burji demonstratives and pronouns. These abstractions do not replace the immediate PHEC reconstructions; they provide a possible account of why accusative exponents are accented and fully voiced while nominatives are typically unaccented and devoiced.

4.1.6 Personal Pronouns

Personal pronouns in Highland East Cushitic distinguish between nominative and accusative cases in all five languages. The comparison isolates recurring pronominal stems and segmentable case exponents, reconstructing paradigmatic structure only where correspondences are stable. Accusative forms (§4.1.6.1) show closer cross-language agreement than nominative forms (§4.1.6.2), especially in the plural, where Burji diverges from the NHEC languages in both stem shape and case-vowel reflexes. Across both paradigms, plural pronouns are typically marked by the plurative ***-n** (§4.1.1.2).

The paradigm distinguishes first, second, and third-person in singular and plural; only the third-person singular shows an obligatory gender contrast. The singular series is largely stable and supports reconstruction to PHEC with

limited ambiguity. The plural series shows extensive generalization (analogical leveling of stems and/or terminal vowels across cells) and paradigm restructuring (reallocation of inherited material across case/person slots).

4.1.6.1 Accusative Pronouns

The accusative (object) pronoun paradigm provides the clearest evidence for reconstructing PHEC case vowels. Singular forms show high cross-language regularity and support reconstruction of the main stems (**ey* 'me', **ke* 'you' (sg), **is-* '3rd person') together with their accusative exponents. Plural accusatives consistently retain the nasal plurative ***-n**, but their additional material varies by language and is treated here as later paradigm restructuring (i.e., language-specific reallocation or extension beyond the shared **-n** base). The generalized forms (i.e., leveled stems) are placed between parentheses. The following lists present the attested reflexes alongside PHEC and PNHEC reconstructions:

First-Person Object

	[103]		[197]	
PHEC	***e-y**	'me'	***n-y**	'us'
Burji	**ʔeé**	'me'	**niísi̯**	'us'
PNHEC	***e-y**	'me'	***n-y**	'us'
Had	**ʔeésẹ̯**	'me'	**neésẹ̯**	'us'
Kam	**ées**	'me'	**nées**	'us'
Sid	**ʔané**	'me'	**ninké**	'us '
Ged	**anee/aneʔe**	'me'	**(noʔooʔo)**	'us'

The first-person accusative pronouns are reconstructed with the **-y** shared by the other first-person pronouns (§4.1.4.2). Much of the diversity among surface cognates reflects morphological combination rather than stem replacement. Burji preserves a stem *ee* (on glide loss rule iii(b), see §2.4 on accent placement). In Hadiyya and Kambaata, *ee* coalesces under combination with the pronominal stem *ise*, yielding *eese* (← **ey* + *ise*). In Sidaama and Gedeo, the same accusative stem combines with the first-person nominative base (*an-*), yielding *anee* (← **an* + *ee*), with **-ʔ-** epenthesis in Gedeo reflecting the treatment of the lost glide (§3.3.4.1).

In the plural [197], note the preservation of the nasal onset ***n** across the family (see §4.1.1.2). Burji, Haddiya, and Kambaata add *ise* after **ny*, yielding *niis* in Burji; the combination of the singular **ey* and the stem *ise* yields Hadiyya and

Kambaata *nees* (← *ni* + *ise*). Sidaama *ninké* includes the additional clitic **=k**, common in the genitive constructions (§§4.1.4.1, 4.1.4.2), and the **-n-** is likely inserted as a placeholder for the dropped glide **y** (§§3.4.2, 2.1.4). The accusative pronoun of Gedeo *noʔooʔo* shows generalization from the nominative counterpart *noʔo* 'we' with additional suffix **-o** preceded by epenthetic **-ʔ-** (§3.3.4.1) in the accusative. Sidaama has maintained only accent contrast between the nominative and accusative. Below is an overview of the singular and plural accusative and the singular possessive, distinguished only by the exponents preceding ***-y**:

	[103]		[197]		[135]	
PHEC	***e-y**	'me'	***n-y**	'us'	***ʔ-y**	'my'
PNHEC	***ey**	'me'	***n-y**	'us'	***ʔy**	'my'

Second-Person Object

The following list summarizes the second-person accusative forms. The focus is on the inherited accusative stem **ke* and its combination with other pronominal material in language-specific compounds:

	[257]		[196]	
PHEC	***ke**	'you' (sg)	***ki-n-**	'you' (pl)
Burji	**ʃeé**	'you' (sg)	**ʃinsi**	'you' (pl)
PNHEC	***ke**	'you' (sg)	**(*kiʔn-)**	'you' (pl)
Had	**keésẹ**	'you' (sg)	**kiʔn-úwwạ**	'you' (pl)
Kam	**kees**	'you' (sg)	**kiʔneʔéeta**	'you' (pl)
Sid	**ʔaté**	'you' (sg)	**kiʔné**	'you' (pl)
Ged	**ateʔe/atee**	'you' (sg)	**(haʔnooʔo)**	'you' (pl)

Second-person singular accusative forms reflect the same layering seen in the first-person: a stable, inherited accusative stem **ke* plus language-specific combinations (often realized as compounds with other pronominal material). Burji continues *ke* with regular palatalization in the relevant environments (§3.2.1). Hadiyya and Kambaata show vowel coalescence under combination with *ise* (*keese*). Sidaama and Gedeo combine *ke* with the second-person nominative base *ati*, producing compounds that are further shaped by regular lenition and epenthesis processes (§§3.2.2, 3.3.4).

The second-person plural accusative base is segmentable as [196] **ki-n-* with **-n** functioning as the pronominal plurative marker across HEC. The PNHEC cognate **kiʔn-* reflects a generalized second-person plural possessive

pronoun [187] **kin-ʔ-* → **kiʔn-* (with the genitive ***-ʔ**, see §4.1.4.1). The remaining NHEC material is language-specific: Hadiyya adds plurative **-uwwa** and Kambaata adds the accusative **-éeta**. Gedeo added the accusative **-o** to the second-person plural nominative stem [104] *haʔno* 'you' (pl) and generalized it to the accusative (*haʔnooʔo*). These extensions pattern with other plural and case morphology discussed in §§4.1.3, 4.1.5.2, and with the morpheme boundary alternations discussed in §§3.3.2, 3.3.4.1.

Third-Person Object

The third-person accusative series isolates the gender-case vowels most clearly. The list is organized to separate the shared stem **is-* from the accusative vowel exponents and plurative ***-n**.

	[151]		**[292]**		**[86]**	
PHEC	***is-e, *is-é**	'her'	***is-o, *is-ó**	'him'	***is-n-**	'them'
Burji	**ʔiʃeé**	'her'	**ʔisi̱**	'him'	**isinoo**	'them'
PNHEC	***is-e**	'her'	***is-ó**	'him'	***is-n-**	'them'
Had	**ʔíse̱**	'her'	**it't'o**	'him'	**(ʔit't'-úwwa̱, it't'u), issʔu**	'them'
Kam	**isé-ta**	'her'	**isú**	'him'	**issoʔóota**	'them'
Sid	**ʔisé**	'her'	**ʔisó**	'him'	**ʔinsá**	'them'
Ged	**isee**	'her'	**isoo, isooʔo**	'him'	**insaʔnee, insaʔné, issaʔné,**	'them'

Third-person accusative pronouns provide the clearest evidence for case-gender vowel exponents. The feminine **is-e* 'her' and masculine **is-o* 'him' share a stem **is-* and contrast only in their terminal vowel, supporting the reconstruction of an accusative gender opposition in PHEC. Hadiyya and Sidaama preserve this opposition most transparently. Kambaata replaces ***-o** with **-ú**, and Burji neutralizes the inherited vowel opposition via analogy with the nominative series, maintaining the accusative-nominative contrast primarily through accentual differences.

The third-person plural accusative form builds on the same stem, with **is-n-* segmentable as 3rd person + plurative ***-n**. Additional suffixes are language-specific and align with the plural extensions already observed in the second-person plural accusative series.

4.1.6.2 Nominative Pronouns

The nominative (subject) pronoun paradigm retains the same person-marking stems as the accusative series, but it replaces accusative case vowels with nominative exponents. The third-person singular is especially informative: *is-* combines with nominative gender vowels (**-i**/**-y** ‘she’, **-u**/**-w** ‘he’) that correspond to the accusative opposition (**-e** vs. **-o**). Plural nominatives consistently retain ***-n** but show extensive generalizations (placed between parentheses), restructuring in stem shape, and addition of suffixes. The following lists of examples present the nominative reflexes with PHEC and PNHEC reconstructions:

First-Person Subject

	[2]		[63]	
PHEC	***an-i**	‘I’	***naʔ-**	‘we’
Burji	**ʔáṇ̩**	‘I’	**naánṵ**	‘we’
PNHEC	***ani**	‘I’	***naʔ-**	‘we’
Had	**ʔáṇ̩**	‘I’	**neésḭ, neese**	‘we’
Kam	**án**	‘I’	**naʔóot**	‘we’
Sid	**ʔani**	‘I’	**(nínke)**	‘we’
Ged	**ani**	‘I’	**noʔo**	‘we’

The first-person singular nominative **an-i* ‘I’ is uniformly attested across HEC. The stem *an-* is stable, and the final **-i** is consistent with the nominative vowel exponent (cf. the case-vowel discussion in §4.1.5.2).

The plural first-person nominative series supports a common base that is related to the first-person accusative **n-y* (§4.1.6.1), but the surface forms largely reflect combination and analogy. Burji attaches additional pronominal material (*an* ‘I’) to the base, while Sidaama and Gedeo generalize a single stem across the nominative and accusative. Sidaama generalizes the accusative stem to the nominative, maintaining the contrast through accent only: *nínke* (nom) vs. *ninké* (acc). Given these generalizations and restructurings, the plural base is best treated as provisional.[180]

[180] There are additional language pair comparisons: Kambaata and Gedeo share the same terminal vowel **-o** before the addition of **-ti** in Kambaata, so prior to the suffixation; Kambaata *naʔ-* is more conservative. Kambaata **-t** and Sidaama **-k** are referenced to the gender exponents discussed in §4.1.5.1. The **-n-** before the **-k** in Sidaama might be the result of the same process of stem **ʔ** + **k** fusion into **kk** and degemination and epenthesis into **-n-k** that occurred with the genitive in Burji and Gedeo (§§4.1.4.1, 4.1.4.4).

Second-Person Subject

The next list presents the second-person nominative series. The singular is expected to reflect the inherited **at-* base with a nominative vowel, while the plural series is evaluated separately because it shows deeper stem divergence between Burji and the NHEC languages.

	[27]		[104]	
PHEC	***at-i**	‘you’ (sg)	***ati-n-**	‘you’ (pl)
Burji	**aʃi**	‘you’ (sg)	**aʃinu**	‘you’ (pl)
PNHEC	***at-i**	‘you’ (sg)	***aʔ-n-**	‘you’ (pl)
Had	**ʔátị̥**	‘you’ (sg)	**(kiʔnúwị̥)**	‘you’ (pl)
Kam	**át**	‘you’ (sg)	**aʔnoʔóot**	‘you’ (pl)
Sid	**ʔáti**	‘you’ (sg)	**(kiʔne)**	‘you’ (pl)
Ged	**ati**	‘you’ (sg)	**haʔno**	‘you’ (pl)

The second-person singular nominative **at-i* ‘you’ (sg) is securely reconstructed. Burji **ʃ** (← ***t**) reflects regular palatalization before high front vowels (or older ***y**; see §3.1.2.1).

The divergent plural bases **ati-n-* (PHEC) and **aʔ-n-* (PNHEC) between Burji and the NHEC languages suggest either early replacement or restructuring. Kambaata and Gedeo share the provisionally reconstructed **aʔ-n-* with additional suffixes. Burji *aʃinu* ‘you’ (pl) extended with the plurative ***-n** also looks like a generalization of the singular counterpart *aʃi* ‘you’ (sg). But unlike the NHEC stems, it has more EC parallels (see Yaaku, Ts’amakko, Dhaasanac). Therefore, second-person nominative **ati-n-* is provisionally reconstructed for PHEC based on retention in Burji, and **aʔ-n-* for PNHEC based on distinctive cognates in Kambaata and Gedeo. The parallel stem *kiʔn-* in Hadiyya and Sidaama (together with the accusative counterpart [196]) is a generalized variant of the second-person plural possessive pronoun [187] **kin-ʔ-* → **kiʔn-* (with the genitive ***-ʔ**, see §4.1.4.1). The comparison below shows the stem generalization between the NHEC accusative (acc), nominative (nom), and possessive (poss) pronouns:

	[196]	acc	[104]	nom	[187]	poss
Had	**(kiʔn-úwwạ)**	‘you’ (pl)	**(kiʔnúwị̥)**	‘you’ (pl)	**kiʔnúwwi**	‘your’ (pl)
Sid	**(kiʔné)**	‘you’ (pl)	**(kiʔne)**	‘you’ (pl)	**kiʔneʔee, kiʔnee**	‘your’ (pl)
Kam	**(kiʔneʔéeta)**	‘you’ (pl)	**aʔnoʔóot**	‘you’ (pl)	**kiʔneʔée**	‘your’ (pl)
Ged	**(haʔnooʔo)**	‘you’ (pl)	**haʔno**	‘you’ (pl)	**(haʔnnoke, haʔnnot’t’e)**	‘your’ (pl)

The parallels between parentheses reflect generalization (leveling) and paradigm restructuring (reallocation across cells) to fit a different paradigm; for example, Gedeo has reshaped the nominative stem and extended it into the accusative and possessive.

Third-Person Subject

The third-person nominative paradigm tests whether the accusative stem *is-* is retained and whether nominative gender vowels (***-i**/***-u**, with glide variants) are preserved. The plural forms are treated as stem + plurative ***-n** with language-specific extensions.

	[6]		[7]		[11]	
PHEC	***is-i, *is-y**	‘she’	***is-u, *is-w**	‘he’	***is-n-**	‘they’
Burji	**ʔíʃi̱**	‘she’	**(ʔísi̱)**	‘he, it’	**ʔisínu̱**	‘they’
PNHEC	***is-i**	‘she’	***is-u**	‘he’	***is-n-**	‘they’
Had	**ʔísi̱**	‘she’	**(isse/issoʔi, it’t’o)**	‘he/she’ (honorific), ‘he’ (subj)	**(it’t’uw), issuw**	‘they’
Kam	**(íse)**	‘she’	**ís/ísu**	‘he’	**issoʔóot**	‘they’
Sid	**(ise)**	‘she’	**(ʔísi)**	‘he’	**ʔínsa**	‘they’
Ged	**(ise)**	‘she’	**(isi)**	‘he’	**(insaʔnne, insaʔne)**	‘they’

The third-person singular nominative forms preserve the inherited case-gender contrast most transparently: **is-i*, **is-y* ‘she’ (feminine) and **is-u*, **is-w* ‘he’ (masculine). The Burji palatalized reflex **ʃ** in the feminine cognate might be a result of **s** + **y** fusion as proposed in (§4.1.5.2 (subsection on Nominative)). The vowel of the case-gender has been leveled across the feminine and masculine, except in Kambaata. Hadiyya again shows the unique glottalized and geminated **t’t’** that it has in the third-person masculine and plural accusative pronouns, and still retains the older reflexes in the honorific pronoun. The **t’t’** is likely a generalization of the third-person singular (m/f) possessive pronoun stem [279] *ʔít’t’i* (via **is=t-ʔ-*) to all the third-person pronouns except the singular feminine nominative and accusative.

The plural **is-n-* builds on the singular with the plurative ***-n**, with retention in Sidaama, addition of a suffix in Kambaata, and generalization of the possessive stem in Gedeo and Hadiyya (§4.1.4.2). The glottal stop **ʔ** in NHEC, and the individual languages restructured the accusative and nominative counterparts heavily, as the following comparison shows:

	[86]	ACC	[11]	NOM
Burji	**isinoo**	‘them’	**ʔisínụ**	‘they’
Had	**(ʔit’t’-úwwạ, it’t’u, issʔu)**	‘them’	**(it’t’uw), issuw**	‘they’
Kam	**issoʔóota**	‘them’	**issoʔóot**	‘they’
Sid	**ʔinsá**	‘them’	**ʔínsa**	‘they’
Ged	**(insaʔnee, insaʔné, issaʔné)**	‘them’	**(insaʔnne, insaʔne)**	‘they’

The terminal vowels of the plurals are not reconstructed because synchronic exponents diverge across languages. Two candidates emerge from terminal vowel quality: ***-uwa** (back vowels) and ***-eye** (front vowels); e.g., PHEC [11] **isn-uwa* ‘they’ (§4.1.1.2). A broader East Cushitic comparison may provide further evidence for recovering such exponents.

For the reconstruction of the possessive pronouns, see §4.1.4.2.

The personal pronoun paradigms support the following PHEC reconstructions: (a) a shared stem **is-* for the third-person across accusative and nominative; (b) case-marking terminal vowels ***-e** (acc), ***-i** (nom) for the feminine; ***-o** (acc), ***-u** (nom) for the masculine; (c) the plurative ***-n** is segmentable in the plural forms; and (d) the genitive ***-ʔ** deducible from the possessive pronouns. The first and second-person paradigms show greater divergence between Burji and NHEC. There is no evidence of case distinction in the third-person plural forms.

Summary of the strategies applied to the pronouns in HEC.

	Generalization (leveling)	Restructuring (reallocation)	Combination (compounding /fusion)	Addition (extensions)
Burji	x	x	x	x
Sid	x	x	x	x
Ged	x	x	x	x
Had	x	x	x	x
Kam	x	x	x	x

Generalization/leveling and paradigm restructuring are especially visible in the plural series and in the parallels in parentheses. Compounding is exemplified in the discussion on 1sg/2sg. Addition is exemplified by language-specific plural and case extensions.

The preceding sections establish the main nominal markers that can be compared across HEC, including number and nominal extensions, the exponents used in genitive constructions, and the primary gender and case markers. Some patterns recur broadly and support reconstruction to higher

levels, while others are restricted to individual languages or to the PNHEC group and therefore constitute candidates for consideration as later innovations.

The next section turns to verb morphology. The verbal system supplies a second set of reconstructable markers, especially in derivational morphology, and it provides many of the triggers for the morpheme boundary processes described in Chapter 3. Reconstructing verbal suffixes and their allomorphy also helps anchor fossilized morphemes in cognates of the reconstructed lexicon discussed in Chapter 2.

4.2 Verb Morphology

This section reconstructs the major components of HEC verb morphology with an emphasis on derivational markers and support verbs. The analysis compares the formal exponents of passive (§4.2.1), causative (§4.2.2), reflexive (§4.2.3), and denominative (§4.2.5) morphology and evaluates their distribution and conditioning across the five languages, in addition to the support verbs (§4.2.4) used in denominative derivation. There is considerable formal overlap between the denominatives and the other markers. Therefore, some of the examples discussed in one section (e.g., the section on Causative -is) may also (or alternatively) fit in the other (e.g., The Causative Denominative).

Where a marker shows regular allomorphy, the relevant phonological or morphological environments are stated explicitly. Verb morphology is directly tied to the boundary processes documented in Chapter 5, since several derivational suffixes trigger assimilation, palatalization, fusion, and epenthesis patterns at the stem-suffix boundary. For this reason, the discussion keeps the reconstruction of each morpheme aligned with the morphophonological evidence and distinguishes inherited markers from later innovations and co-morphologies (contact-induced).

4.2.1 The Passive

The passive suffix **-am** is attested in Burji (Tesfaye Baye, 2015, p. 181), Hadiyya (Ritter, 2007, p. 6), Kambaata (Treis, 2008, p. 357), Sidaama (Yri & Pepper, 2019; Kawachi, 2007, p. 342), and **-em** is attested in Gedeo

(Wedekind, 2008, p. 2; Gasparini, 1995, p. 97). Therefore, PHEC ***-am** is securely proposed.[181]

	ø/-ʔ		**-am**	
Had	**dun-**	'to leak, to pour'	**dun-am-**	'to be poured, to leak out'
Kam	**dun-**	'spill'	**dun-am-**	'be spilled, be poured'
Sid	**dun-**	'pour out'	**duʔn-am-**	'be poured out'
Ged	**dun-**	'pour out'	**dun-em-**	'flow out, be poured'
Burji	**dunʔ-**	'to poke'	**-**	-

Sidaama *duʔnam-* 'be poured out' (pass), from the derived stem *duʔn-* 'pour out', is identified as autobenefactive in Yri & Pepper (2019). Hudson (1989, p. 89) has Sidaama *dun-am-* 'leak (of e.g. water)' from the underived stem. Sidaama *duʔn-* 'pour out' is a parallel to Burji *dunʔ-* 'to poke'. Alemu Banta's (2016) glossing of the Kambaata passive *dun-am-* 'spilling out, pouring out' is inaccurate and should be 'be spilled/poured out' as in the rest of NHEC.

Gedeo **-e** is also found in the middle variant **-ed'**. Wedekind (1980, p. 149) considers such a correspondence of **a** and **e** to be related to stress placement,[182] e.g., Sidaama *galado* and Gedeo *galedo* 'monkey'.

4.2.2 The Causative

The PHEC causative marker is ***-s**. It is retained as **-s** following single stem-final sonorants, is preceded by an epenthetic **-i-** as **-is** following stem-final geminates and clusters, and in PNHEC, is fused and geminated as ***s-s** following single stem-final obstruents (see §3.1.2.2 for related processes). A palatalized variant **-ʃ** is found in Kambaata, Sidaama, and Gedeo, and an affricated **c** is found in Gedeo.[183]

[181] Data sources: Hadiyya (Ritter, 2007, p. 34), Kambaata (Alemu Banta, 2016; Treis, p.c.), Sidaama (Yri & Pepper, 2019), Gedeo (Wedekind, 2008, p. 48), Burji (Getahun Addo, 2017, p. 224). See also Hudson (1989, p. 89; 2006, p. 296).

[182] "For Sidamo, the basic stress rules can be summarized from Hudson's paper (1969, 13f.): (a) The penultimate syllable of the morpheme is stressed, (b) but some affixes and some verb forms demand word final stress. For Gedeo, the basic rules appear to be the opposite: (a) The final syllable of the word stem is stressed, (b) but most suffixes and some noun and verb forms demand penultimate stress". (Wedekind, 1980, p. 139)

[183] Data sources: Kambaata (Treis, 2022), Gedeo (Eyob Kelemework, 2015), Sidaama (Kawachi, 2007), Hadiyya (Tadesse Sibamo, 2015; Ritter, 2007), Burji (Tesfaye Baye, 2015). Only a few examples are from the database. Where translations are missing, it means they are not provided in the source.

4.2.2.1 Causative *-s

The causative **-s** is less frequently attested than **-is**.

Had	*t'an-* 'do, perform' → *t'an-s-* 'to enable somebody to do something'
	mar- 'to go' → mas-s- 'to take, to let go'
Kam	*zaaz-* 'flow' → *zaas-s-* 'make flow'
	sagab- 'become thirsty' → *sagas-s-* 'make thirsty'
Sid	*makkal-* 'to become crazy/mad' → *makkal-s-*
	mool- 'to become dry' → *mool-s-* / *moo-ʃʃ-*
Burji	*hiik'-* 'close' → *hiisk'-* (with metathesis)
	goob- 'be brave' → *gooss-*, *goobis-*

In Kambaata and Sidaama, stem-final obstruents sometimes fully assimilate to the attached causative **-s** (§3.1.1.3). The full assimilation of labials and alveolars has been preceded by a metathesis process (**C+s** → **sC**) in Burji, which is still retained before velars. Gedeo is the only HEC language that does not have an allomorph **-s** without epenthetic **-i-**.

4.2.2.2 Causative -is

The frequently and widely attested variant **-is** of the causative comprises the epenthetic **-i-** and the causative **-s**.

Had	*holl-* 'chase'→ *holl-is-*
	ʔaanʃ- 'wash' → *ʔaanʃ-is-*
Kam	*birr-* 'stop (of rain)' → *birr-is-* 'take shelter (from rain)'
	c'ung- 'drink the last drop (of a drink)' → *c'ung-is-* 'empty (a container) to the last drop'
Sid	*c'oomm-* 'to taste good' → *c'oomm-i-s-* / *c'oomm-i-ʃ-*
Ged	*fiʃʃ-* 'smell' → *fiʃʃ-i-s-* / *fiʃʃ-i-ʃ-*, *darʃ-* 'swell' (intr) → *darʃ-i-s-* / *darʃ-i-ʃ-* 'swell' (tr)
Burji	*damm-* 'to pass or spend the night' → *dammis-* 'to allow someone to pass or spend night in one's home'

The formal and functional overlap with the derived verb **is* 'to make' also contributes to its prominence across all the HEC languages, as discussed in §3.1.1.3. As already seen in the Sidaama and Gedeo examples above, there is a palatalized variant **ʃ**, which shows free variation with **s**.

4.2.2.3 Causative -ʃ

In Kambaata, Sidaama, and Gedeo, the causative ***s** has been recently palatalized into **ʃ** (§3.1.2.2).

Kam	*ʃiil-* 'melt' → *ʃiiʃ-ʃ-*
Sid	*bub-* 'to burn' → *buʃ-ʃ-*
	duub- 'to become satisfied' → *duuw-i-s-* / *duuw-i-ʃ-* / *duuʃ-ʃ-*
Ged	*utal-* 'jump' → *utal-c-*
	leell- 'appear, be seen' → *leell-i-ʃ-* 'show'

The innovated palatalized **-ʃ** occupies all the environments where the allomorphs of the alveolar **-s** occur, but it is most prevalent after sonorants, particularly **l**, as in Kambaata. The environment after the lateral **l** is also where Gedeo has innovated the affricated variant **c** of the palatalized **ʃ**, e.g., *utal-* 'jump' → *utal-c-*.

4.2.2.4 Double Causative

The double causative (caus1 + caus2) includes a causative and complex form (**-s-y-i-s**) comprising the causative **-s** (caus1) followed by another causative but preceded by a support verb **y** and epenthetic **-i-** yielding **-siis**.

	Stem		**Single causative**	**Double causative**
Had	**ʔag-**	'drink'	→ **ʔag-is-**	**ʔag-isiis-**
Kam	**waal-**	'come'	→ **waaʃ-ʃ- 'bring'**	**waaʃ-ʃ-isiis-**
Sid	**t'orʃ-**	'to pour'	**-**	→ **t'orʃ-isiis- / t'orʃ-iʃiiʃ-**
Ged	**ful-**	'get out'	→ **ful-c-**	**ful-ciis**
	raʔ-	'ripen'	→ **raʔ-is-**	**raʔ-isiis- / raʔ-iʃiiʃ-**
	ib-	'fall'	→ **ib-is-**	**ib-isiis-, ib-is-isiis-**
Burji	**ʔit-**	'eat'	→ **ʔit-is-**	**ʔit-is-iss-**[184]

All the HEC languages have a double causative with the non-palatalized **-siis-**, while multiple causatives of the form **-s-isiis-** are found in Gedeo, Kambaata, and Burji only. Gedeo, Kambaata, and Sidaama have instances of the palatalized **ʃ** (and Gedeo **c**) (caus1) and alveolar **s** (caus2), e.g., **-ʃ-isiis-** in the double/multiple causative formations. Gedeo has a variant **c** that occurs after stem-final **l**, as in *ful-* 'get out' → *ful-c-*, *ful-ciis*. Gedeo and Sidaama have optional alternation between the palatalized harmonized **-ʃiiʃ-** and non-palatalized alveolar **-siis-** allomorphs. Burji and Hadiyya have a

[184] Kambaata also has *it-* 'eat, rip off' → *it-is-* 'feed, offer (food)', *it-siis-* 'cause to be ripped off'.

geminate **ss** in caus2 via the addition and fusion of the verb **y* 'to say, do, be' and the caus2 **-s**, i.e., **-siss**. For example, in Hadiyya, the double causative of *kiʔ-is-* (caus1) 'to make somebody stand up, to wake up' is *kiʔ-isiss-* (caus1 and caus2) 'to wake somebody up' from the stem *kiʔ-* 'to stand up, to wake up'. Otherwise, Hadiyya does not share most of the other double causative formations.

A summary of the allomorphs shared by two or more HEC languages is provided below:

	Single causative	i/y-causative	Double	Multiple
Had, Kam, Sid, Burji	**-s**			
Kam, Sid, Ged	**-ʃ**			
all HEC	**……**	**-is**		
Kam, Sid, Ged	**……**	**-iʃ**		
all HEC	**……**		**-siis**	
Burji, Had	**……**		**-siss**	
Ged, Sid	**……**		**-ʃiiʃ**	
Ged, Kam, Burji	**……..**			**-sisiis**
Ged, Sid, Kam	**……..**			**-ʃisiis**

Multiple causatives are allowed in all HEC languages. However, the palatal variant **-ʃ** of the causative suffix in Kambaata, Sidaama, and Gedeo never occurs as a second causative, i.e., it is not attested as caus2 following and **-s** causative. The reconstructable forms are three allomorphs of the single causative – ***-s**, ***s-s** (PNHEC), and ***-is** – and one double causative ***-siis**. However, some of the other listed double/multiple causative forms might have also existed at the proto-levels.

4.2.3 The Reflexive

The HEC literature variably labels the same exponent as reflexive, middle, or autobenefactive. The use of "reflexive" here as a cover term broadly includes middle and autobenefactive functions (§1.5.2), which are formally and functionally indistinguishable synchronically across HEC languages (§4.2.3.1). In the HEC languages, these functions are expressed by a single morphological marker (**-ʔ** and its allomorphs), making them formally indistinguishable. Glosses in this section occasionally specify the function where it has been provided in the source, but the functional boundaries are fluid, language-specific, and beyond the scope of this study. Formally, the historical proposal assumes that the middle and autobenefactive functions are

served using a reflexive marker ***ʔ** and the combination of ***ʔ** and ***t**, ***k** (***-t-ʔ-**, ***-k-ʔ-** → ***-t't'**, ***-k'k'**).

This section is organized as follows: §4.2.3.1 presents the basic reflexive ***-ʔ**, documenting its distribution, its alternations conditioning (metathesis after sonorants, fusion after obstruents), and the theoretical justification for reconstructing **ʔ** rather than the commonly posited **d'**. §4.2.3.2 examines the complex reflexive **-ak'k'** found primarily in Kambaata and Hadiyya. §4.2.3.3 analyzes the reflexive **-it't'** in Hadiyya and Kambaata as the historical fusion of (agentive) **-t** and reflexive **-ʔ**, presenting synchronic evidence and discussing the limits of stem comparability. §4.2.3.4 addresses the **-Vd'/-d'** variant in Burji, Gedeo, and Sidaama, attributing it to Oromo contact influence and distinguishing it from the inherited PHEC system.

4.2.3.1 Reflexive *-ʔ

The suffix **-ʔ** is found as a reflexive in all HEC languages: Burji (Tesfaye Baye, 2015, p. 183), Hadiyya (Tadesse Sibamo, 2015, p. 131), Kambaata (Treis, 2008, p. 217), Sidaama (Kawachi, 2007, p. 335), and Gedeo (Eyob Kelemework, 2015, p. 190). The productivity and antiquity of this suffix are evident from its systematic participation in morpheme boundary alternation processes: fusion following obstruent-final stems (§3.3.1) and metathesis following sonorant-final stems (§3.3.2). These processes, shared across NHEC languages, indicate that ***-ʔ** has been a stable component of the HEC verbal system since P(N)HEC.

In NHEC, the morpheme boundary alternations of the reflexive suffix **-ʔ** are conditioned by whether it attaches to a stem ending in sonorant or obstruent consonants, i.e., metathesis following stem-final sonorants (§3.3.2) and fusion following stem-final obstruents (§3.3.1).

The following examples illustrate the sonorant-final environment. The presentation contrasts NHEC forms showing metathesis with corresponding stems in Burji, Kambaata, and Gedeo, where metathesis is absent:[185]

[185] Data sources (additional data): Gedeo (Wedekind, 2008, p. 19), Kambaata (Alemu Banta, 2016), Burji (Getahun Addo, 2017, p. 306).

			[36]				[120]	
Burji	**gam-**	‘to bite’	**-**	-	**fil-**	‘to comb’	**silʔ-**	‘to comb for oneself’
Had	**-**	**-**	**gaʔm-**	‘to bite’	**-**	**-**	**-**	-
Kam	**gam-**	‘cause bellyache’	**gaʔm-**	‘to bite’	**-**	**-**	**fiil-**	‘unravel’
Sid	**-**	**-**	**gaʔm-**	‘to bite’	**-**	**-**	**-**	-
Ged	**-**	**-**	**gaʔm-**	‘to bite’	**fil-**	‘comb’	**fiʔl**	‘to comb’

The sets show that metathesis is characteristic of the NHEC reflexive system, whereas the corresponding cognate [120] in Burji does not show metathesis. Where no underived stem is attested, the most economical interpretation is fossilization rather than productive derivation.[186]

When **-ʔ** follows a stem-final obstruent, glottalization occurs through fusion (§3.3.1). This process results in the root-final consonant becoming glottalized and geminated (**C’C’**).

b, d/t, k/g, ʃ/j/z/s + ʔ → p’p’, t’t’, k’k’, c’c’

The following examples give representative stem-reflexive pairs showing the regular fusion outcome (**C + ʔ → C’C’**) across NHEC:[187]

	Stem		**Reflexive**	
Had	**amad-**	‘to catch’	**amat’t’-**	‘to hold some parts of the body’
Kam	**kuʃu**	‘choke’	**kuc’c’u**	‘choke, cough’
Sid	**daak-**	‘to grind’	**daak’k’-**	‘to grind (for oneself)’
Ged	**dib-**	‘prepare ensete food’	**dip’p’-**	‘prepare food for oneself’
cf. Burji	**ʃaf-**	‘carve’	**ʃad’d’-**	‘to carve for oneself’

As shown above, Burji has a unique development **f+d’, b+d’** → **d’d**’ as seen in *ʃad’d’-* ‘carve for oneself’ with exceptions, e.g., *gaaf-* ‘grind’ → *aafd’* ‘grind

[186] Burji shows cases of free variation between the allomorphs **ʔ** and **d’**, and Sasse (1982, p. 18) notes that Burji **d’** and **ʔ** are dropped before consonants with oncomitant lengthening of the preceding vowel.
[187] Data sources: Gedeo (Wedekind, 2008, p. 58), Sidaama (Kawachi, 2007, pp. 57–58, 35; Indrias Xa’miso et al., 2007, p. 101), Hadiyya (Tadesse Sibamo, 2015, p. 133; Ritter, 2007, p. 7), and Kambaata (Treis, 2008, pp. 243, 269, 242, 267, Treis, p.c.), Burji (Getahun Addo, 2017, p. 777).

for oneself' (Tesfaye Baye, 2015, pp. 182–183) and shows no evidence of a fusion process similar to that found in the NHEC languages.

This analysis differs from all the existing HEC literature (at least on Gedeo and Sidaama) in proposing the glottal stop **ʔ** as the sole variant responsible for fusion and glottalization instead of the commonly posited **d'**. Five arguments converge to support reconstructing ***-ʔ**:

(a) Parsimony: There is no reason for postulating **d'** with subsequent change to **ʔ**, then metathesis and fusion, when an established allomorph **ʔ** already exists in the paradigm. Occam's Razor favors projecting the attested **ʔ** directly to the proto-level rather than positing an intermediate ***d'**.

(b) Traces of assimilation: The hypothetical allomorph ***d'** lacks the expected traces of partial assimilation it should have left on adjacent **m**, in the form of **nd'**, before changing to **ʔ**, that other alveolars have left (§3.1.1.4). Comparative evidence from NHEC shows consistent alveolar assimilation (e.g., ***-m** → **-nd-** / __ **d**); the absence of **nd'** patterns undermines the ***d'** hypothesis.

(c) Phonetic plausibility: The fusion of the voiceless glottal stop **ʔ** with both voiced and voiceless obstruents is phonetically more plausible in terms of yielding regular voiceless glottalized geminates **C'C'** than fusion with the voiced implosive **d'** (§3.3.1). Therefore, it is a fusion of **ʔr** that yields the expected glottalized **d'(d')** in Gedeo and Sidaama (§3.3.1.5).

(d) Phonemic status: There is limited evidence to establish **d'** as a stable phoneme in minimal pairs in the relevant HEC languages (Tesfaye Baye, 2015, p. 27; Kawachi, 2007, p. 35). If **d'** were a core proto-phoneme, one would expect robust paradigmatic contrast; instead, it shows restricted distribution and contextual variation.

(e) Distribution: The consonant **d'** is restricted to Burji, Sidaama, and Gedeo, i.e., it is not reconstructable to PHEC or PNHEC. Its presence is attributed to alternation process (**ʔr** → **d'**) documented in §3.3.1.5 and Oromo contact influence (§4.2.3.4). This geographically/historically restricted pattern argues against a PHEC origin.

However, suppose there is such an alternation **d' ~ ʔ**. In that case, it must be confined to the environment following a sonorant and must have been

innovated individually due to the foreign nature of the borrowed sound **d'**. Synchronically, **-ʔ** is formally confused with the variant **-d'** because they are functionally indistinguishable (middle ~ reflexive ~ autobenefactive, where those terms are used equally to refer to form and function).

4.2.3.2 Reflexive -ak'k'

The reflexive **-ak'k'** is found in Kambaata (Treis, 2022, p. 17; Treis, 2008, p. 217; Hudson, 1976, p. 272) and Hadiyya (Tadesse Sibamo, 2015, p. 134; Sim, 1991, pp. 54–55, fn. 6). Following is a representative list of Kambaata and Hadiyya examples, organized by stem and the derived reflexive form:[188]

	Stem			**Reflexive**
Kam	**eeb-**	'to bring'	→	**eeb-ak'k'-**
	maat'-	'to hide'	→	**maat'-ak'k'-**
Had	**ʔeeb-**	'to bring/marry'	→	**ʔeeb-ak'k'-**
	mooʔ-	'see'	→	**mooʔ-ak'k'-**

The distribution of **-ak'k'** is currently limited to Kambaata and Hadiyya. Based on the present evidence, this pattern is best treated as a shared innovation (see §4.2.5.5) rather than a proto-level reflexive exponent.

Interestingly, older literature does not report Hadiyya **-ak'k'**. Plazikowsky-Brauner (1957, p. 29) records a reflexive **-ak** in Hadiyya, e.g., *uw-ak* 'to give for oneself'. Hayward (1984, p. 90) considers those Kambaata and Hadiyya forms with **-ak'k'** to be innovations. The older description of **-ak** for Hadiyya may reflect underdocumentation, variation, or different segmentation in the earlier sources. However, the question would remain: Is Hadiyya **-ak'k'** a recently introduced loan via contact with Kambaata, or simply unrecorded in the previous studies? For the morphophonological behavior of the reflexive **ʔ** in fusion processes with obstruents and metathesis with sonorants, see §§3.3.1, 3.3.2.

4.2.3.3 Reflexive -it't'

Hadiyya and Kambaata show synchronic evidence of **-it't'** used in middle forms of words ending in consonants other than **t**. The pattern can be analyzed as the historical fusion of (agentive) **-t** and reflexive **-ʔ**, yielding a glottalized geminate outcome. The sources for these languages do not recognize this

[188] Data sources: Kambaata (Treis, 2022, p. 17; Treis, 2008, p. 193) and Hadiyya (Tadesse Sibamo, 2015, p. 134).

ending as a suffix, so this section presents key examples that motivate segmenting **-it't'** as a complex reflexive ending.

The following examples provide comparisons between reflexive forms in **-it't'** and their most plausible underived stems, illustrating both the complexity of the reflexive ending and the limits of stem comparability in some items:[189]

	Reflexive -it't'		**Stem**	
Had	**uull-it't'-**	'to erect, to put an end' (tr)	**uull-**	'to stand (up), to cease, to stop, to take place (intr)'
			uull-is-	'to stop (tr), to make cease, to make stop (intr)'
Kam	**uurr-it't'-**	'keep still/away'	**uurr-**	'stand up, take place'
			uurr-is-	'stand, erect, stop'

This example shows that **-it't'** occupies the same suffixal slot on the stem as the causative counterpart **-is**.

	Reflexive -it't'		**Stem**	
Had	**k'iʃ-it'-**	'to save, to be kind, to grant forgiveness' (intr/tr)	**k'iiʃ-**	'to make a fence around something, to enclose' (intr/tr)
Kam	**k'iʃ-it't'-**	'have mercy'	**-**	-

While Kambaata shows the regular reflex **-it't'**, Hadiyya has a degeminated **-it'** in this example. The same applies to the following example:

	Reflexive -it't'		**Stem**	
Had	**fir-iit'-**	'to gallop' (intr)	**fir-**	'to go out, to exit, to come out, to get out, to climb, to ascend, to pass' (intr)
Kam	**fir-iit'-**	'gallop (of a horse)'	**firr y-**	'jump in different directions'
			firiic'c'a	'gallop' (m)

This example exhibits that the lexical comparability between derived and underived stems varies across items. The vowels on the reflexive can be interpreted as **y** + **-it'** → **-iit'** as proven by the corresponding Kambaata compound verb stem *firr y-* 'jump in different directions' (see §4.2.4.1). The

[189] The data used in this section comes from Hudson (1989), Ritter (2007), Tadesse Sibamo (2015), Treis (p.c.), and Treis (2008).

same applies to the corresponding derived Kambaata *firiic'c'a* 'gallop' (← **y + -is + ʔ**).

As most examples show, this complex variant of the reflexive is usually preceded by epenthesis **-i-t't'** (§3.4.1). Although the first example is straightforward, not every pair securely reflects a transparent derivational relationship, so individual etymologies should be treated cautiously. In other words, some of the above stems might show distant or no semantic relation to the corresponding reflexive marked stem, e.g., the reflexive stems Hadiyya *fir-ii-t'-* (intr) 'to gallop', Kambaata *fir-ii-t'-* (intr) 'gallop (of a horse)' and the corresponding Hadiyya *fir-* (intr) 'to go out, to exit, to come out, to get out, to climb, to ascend, to pass', Kambaata *firr y-* 'jump in different directions'.

Such a form of reflexive marking has limited attestations in Kambaata, and most of the evidence comes from Hadiyya. Below are further examples from Hadiyya:

Reflexive -it'(t')		**Stem**	
haabb-it't'-	'to comb one's hair'	**haabbis-**	'to comb'
c'aakk-it't'-	'to light oneself'	**c'aakk-**	'to light'
ʔap'p'-it't-	'hit oneself'	**ʔap'p'is-**	'to hit'
huuʃ-it't'-	'miss'	**huuʃ-**	'miss, make mistake'
iibb-it't'-	'to get warm'	**iibb-i-s-**	'to heat, to warm (up)'
		iibb-	'to be hot'

Taken together, the Hadiyya data suggest that **-it'(t')** is at least partly productive and interacts with the same morpheme boundary alternations pressure discussed for ***ʔ** (e.g., avoidance of marked consonant clusters via epenthesis and fusion) (see §4.2.5.5).[190]

In *ʔap'p'-it't'-* (refl) 'hit oneself', in addition to the fused Hadiyya **t't'**, multiple fusion processes are observed in the cognate set provided in Hudson (1989, p. 43) (sometimes on the same stem):

Sid		**kibb-** 'hit with horns', **kibb-am-** 'hit with horns (of rams)'
Had	**bb + ʔ → p'p'**	**ap'p'-is-** 'collide (with)'
	t + ʔ → t't'	**ap'p'it't'-**, **ap'p'it't'-am-** 'collide (with)'
Kam	**s/ʃ + ʔ → c'c'**	**k'oc'c'-**, **k'oc'c'-ak'k'-am-** 'collide (with)'

[190] Th stems with **-it't'** are not consistently intransitive, as one would expect in the reflexive derivation, e.g., *sas-s-iit't'-* 'to make something cooked', *sar-* 'to cook, to bake', and *sigg-it't'-* 'to cool down' (tr), *sigg-* 'to be(come) cool/cold' (intr). Some of these derivations might involve a different process, e.g., a reflexive denominative (§4.2.5.5) in the second example, where the semantics are not always transparent.

The table indicates that glottalization outcomes in this set are not limited to a single consonant class and may involve multiple morphophonological pathways (including fusion and reanalysis), sometimes within the same lexical stem. This variability motivates the discussion below of competing segmentations and historical sources.

Tadesse Sibamo (2015) does not provide any explanation except that "in the process of middle (autobenefactive) stem formation the root final obstruents undergo glottalization. i.e., … *t'* to *tt*'", although the above roots do not end in the claimed **t'** (except another example *t'oreet'- '*compete' → *t'oreet't'-* (refl)).

Kambaata also has *hujat't'-* (mid) 'work for one's own benefit', *hujat-* (tr) 'work', and *huj-íta* (f.acc) 'work', which is comparable with Oromo *hòjjed'd'-à* 'work, make', *hòjjéd'd'a* 'I work' (Lloret, 1988, p. 188). However, the question is whether Kambaata **-t't'** has been introduced recently via contact with Hadiyya or has simply not been recorded in previous studies.

4.2.3.4 Reflexive -Vd', -d'

Parallel to the unique reflexives **-ak'k'** / **-it't'** in Hadiyya and Kambaata, Sidaama, Gedeo, and Burji have a reflexive variant with the consonant **-d'**. Burji allomorphs of this variant include **-ad'** and **-d'** (Tesfaye Baye, 2015, pp. 181–182), Gedeo has **-ed'** (Wedekind, 2008, p. 2; Eyob Kelemework, 2015, p. 190), and Sidaama has **-i-d'** (Kawachi, 2007, p. 335; Anbessa Teferra, 2014, p. 50; Yri & Pepper, 2019). The following are the main allomorphs (**-d'**, **-ad'**, **-ed'**, **-i-d'**) illustrated with representative stems and derived reflexive forms in Burji, Gedeo, and Sidaama:[191]

		Stem			**Reflexive**
Burji	**-d'**	**gaaf-**	'grind'	→	**gaafd'**
	-ad'	**daah-**	'swim'	→	**daah-ad'-**
Gedeo	**-ed'**	**erg-**	'send'	→	**erg-ed'-**
Sidaama	**-i-d'**	**abb-**	'to bring'	→	**abb-i-d'-**

The prominence of the variant **-Vd'** in descriptive grammars of Burji and Gedeo (Eyob Kelemework, 2015, p. 71; Tesfaye Baye, 2015, p. 183) may reflect the fact that this variant can attach to stems without triggering epenthesis, fusion, or metathesis, making it morphotactically simpler than the NHEC pattern. In Oromo, the first-person middle is marked by **-ad'd'** (Ali &

[191] Data sources: Burji (Tesfaye Baye, 2015, p. 182), Gedeo (Eyob Kelemework, 2015, p. 192), and Sidaama (Kawachi, 2007, p. 42).

Zaborski, 1990, p. 22), **-ed'd'** (Orma) (Stroomer, 1987, p. 153), and **-d'd'** (Lloret, 1988, p. 189). Gemination aside, the Oromo variants appear to mirror those of Burji **-ad'** and **-d'** and Gedeo **-ed'** in both consonants and vowels.[192] Despite the existence of the Oromo variant with mid-vowel **e**, it is not uncommon for Gedeo to raise **a** to **e** in suffixes; for instance, the passive ***-am** became Gedeo **-em** (§§2.2.1.1, 4.2.1). That Sidaama **-i-d'** is preceded by epenthetic **-i-** to avoid a **CCC** sequence when added to a stem reveals the underlying quantity of **d'** (← **ʔr**). As argued in §4.2.3.1, the variant **-d'** is formally and historically different from **-ʔ**.

Sidaama and Gedeo additionally suggest an alternation between **ʔr** and a glottalized **d'** (often with unstable quantity **d'd' ~ d'**) in the reflexive. In line with §3.3.1.5, this pattern can be interpreted as a comparatively recent, contact-induced reanalysis of the sequence **ʔ+r** (expected under sonorant + **ʔ** metathesis) as a glottalized **d'(d')**, which also helps explain why quantity is least stable here. Unlike the other non-pulmonic consonants, the borrowed **-d'** is hardly found geminated. Additionally, Plazikowsky-Brauner (1957, p. 29) argues that the reflexive **d'** in Sidaama results from Oromo influence, a hypothesis that aligns with the linguistic history and context of Oromo-Sidaama contact in the region. Therefore, Burji **-ad'** and **-d'** are entirely from Oromo, while the Gedeo and Sidaama **-Vd'** variants are from a morpheme boundary alternation process induced by contact with Oromo (§3.3.1.5).[193]

The Burji forms are argued to be direct loans because Burji does not participate in the fusion process responsible for introducing **d'(d')** in the NHEC languages. Burji still shows cases of free variation between the borrowed reflexive **-d'** and the retained **-ʔ**, e.g., *kanʔi*, *kand'i* 'here it is!'; *sil-*, *silʔ-* / *sild'-* 'to comb'; and *kurʔ-*, *kar-d'-* 'decide'.[194]

Lloret's (1988) analysis of Oromo provides complementary evidence for understanding the middle system in EC. She proposes that the seemingly

[192] The degemination itself could be proof of the borrowed origin of the non-pulmonic **d'** in HEC, since all other non-pulmonic consonants can occur as geminated and in the **nd'** cluster. Likewise, if one accepts the **d'** variant for Sidaama and Gedeo, its purported change into a glottal stop is further proof of its external origin.

[193] According to Wedekind (1980, p. 134), Cerulli equates the Gedeo implosive **ɗ** with the geminated glottalized **t't'** (1929, p. 235). However, Wedekind argues that it is only true on the morpho-phonemic level.

[194] Such free variation between **d'** and **ʔ** following alveolar sonorants is not confined to verbs in Burji, e.g., *k'anʔ-i*, *k'and'-i* 'clitoris'. A similar **ʔ** and **d'** alternation before alveolars is found in the South Omotic languages, e.g., Hamar *guldánti ~ gulʔánti* 'belly button', Ari *gulʔa* 'navel' (Petrollino, 2016, pp. 16, 274).

geminated **d'd'** reflexive marker in Oromo derives from an underlying /**ʔt**/ sequence, where the /**ʔ**/ has no underlying weight. She argues that it explains the morphophonological alternation in Oromo between **d'd' ~ t** without requiring compensatory lengthening: "[W]eightless /ʔ/ is only syllabified when it merges with a following consonant, resulting in [d'd']. If it does not merge, it simply deletes without compensatory lengthening" (Lloret, 1988, p. 198). This analysis, though admittedly abstract, provides an explanation for the synchronic patterns observed in Oromo.

In the broader East Cushitic context, there has been debate regarding the historical development of these middle/reflexive markers. Hayward (1984) proposes that the Proto-East Cushitic system alternated between ***ɗ (= *d')** and ***t** in its middle voice paradigm. This alternation underwent different patterns of leveling in various branches: "[I]n the Highland Eastern Cushitic languages... levelling also took place in the suffixed middle root extension, but in favour of ***ɗ**" (Hayward, 1984, p. 84). This contrasts with the PNEC branch, which leveled toward ***t** instead.

4.2.3.5 Summary and Reconstruction

The comparative evidence presented in §§4.2.3.1-4.2.3.4 supports the following reconstruction of the HEC reflexive system:

> Core PHEC marker: ***-ʔ** is reconstructable to Proto-HEC based on its HEC-wide distribution (attested in all five languages), regular NHEC morpheme boundary alternation (metathesis following sonorant-final stems, fusion following obstruent-final stems), phonological plausibility (voiceless **ʔ** fusing with voiced/voiceless obstruents → voiceless glottalized geminates **C'C'**), and absence of the traces of assimilation (***md'** → ***nd'**) that would be expected if **d'** was the allomorph in the metathesis (§3.1.1.4).
>
> PNHEC innovations: metathesis of sonorant + **ʔ** → **ʔC**, fusion of obstruent + **ʔ** → **C'C'**, and the complex forms **-ak'k'/-it't'** limited to Kambaata/Hadiyya; plausibly analyzable as **-Vk + -ʔ** → **-Vk'k'** / **-Vt + -ʔ** → **-Vt't'** with subsequent reanalysis.
>
> Secondary developments: **-Vd'/-d'** variants in Burji, Gedeo, and Sidaama are attributed to direct borrowing from Oromo (Burji **-ad'/-d'**). There has been contact-induced reanalysis of the sequence **ʔr** as **d'** (Gedeo/Sidaama **-ed'/-i-d'**), hence it is always preceded by a vowel.

> Thus, the phonemic status of **d’** is that it is not reconstructible to PHEC or PNHEC, but limited to individual languages.
>
> Theoretical innovation: This analysis differs from existing HEC literature (Hudson, 1976; Kawachi, 2007; Kelemework, 2015) in proposing **ʔ** as the sole PHEC variant responsible for fusion/glottalization, rather than alternating **d’ ~ ʔ**. The evidence strongly favors projecting ***ʔ** to PHEC, with **d’** emerging later through contact and morphophonological reanalysis.

The next sections (§§4.2.4–4.2.5) examine the support verbs and denominative derivation, where the verbal suffixes (§§4.2.1–4.2.3) participate in the derivation process as part of the denominative suffix attached to nominal stems.

4.2.4 Support Verbs

This section examines two PHEC support verbs: [425] **ʔ* ‘to do’ and [388] **y* ‘to say, do, be’. They differ fundamentally in their inherent transitivity and combine with nominal stems to form compound predicates. Unlike the grammaticalized verbalizer ***ʔ** examined in §4.2.5, these verbs retain some autonomous lexical capacity while participating in derivational structures (see Hudson, 1976, p. 273; Crass, 2007, p. 30; Treis, 2024, p. 165). The transitivity distinction (§4.2.4.3) determines which causative suffix (**ass-** vs. **is-**) a derived stem selects, a pattern that extends through the valency-changing morphology described in §§4.2.1–4.2.3. First, §4.2.4.1 presents the intransitive support verb **y*; §4.2.4.2 then presents the transitive support verb **ʔ*; then, §4.2.4.3 demonstrates their stem-combining patterns across HEC languages.

PHEC [425] **ʔ* ‘to do’, with its derived causative **ass-* (← **ʔs-*), and PHEC [388] **y* ‘to say, do, be’, with its derived causative **is-* (← **y-s*), appear in all HEC languages as independent verbs.

Following Treis (2024), the term “support verb” is used here as equivalent to the term “light verb” of the wider literature. The term “verbalizer” is reserved for the grammaticalized reflex of support verbs when functioning as derivational morphemes on nominal stems. Support verbs in PHEC function simultaneously as independent lexical verbs and as grammatical verbalizers, enabling category change from nominal to verbal stems. This dual function establishes the foundation for the grammaticalization pathway documented in §4.2.5, where support verb constructions evolve from simple denominatives

(§§4.2.5.1–4.2.5.2) through layered combinations (§§4.2.5.3–4.2.5.7) to generalized valency-changing morphology (§§4.2.5.8–4.2.5.9).

The following subsections present each support verb first as an independent lexical item (with cognates demonstrating Proto-East Cushitic inheritance), then examine its grammaticalized function as a verbalizer in denominative derivation. Understanding these dual capacities is essential for interpreting the reanalysis framework in §4.2.5.9, where traditionally analyzed "epenthetic" vowels (**-i-**, **-a-**) are shown to be grammatical reflexes of support verbs **y* and **ʔ*, respectively.

4.2.4.1 The Verb **y* 'to say, do, be'

The PHEC verb [388] **y* 'to say, do, be' is found in all the HEC languages and has a derived causative counterpart **is-* (← **ys-*) 'to make'. The following set presents the stem and causative forms of **y* across the HEC languages. The causative variants show the morpheme boundary alternant after adding the causative suffix ***-s** to the root **y*.

	Stem		**Causative**	
PHEC	***y**	'to say, do, be'	***ys-** → ***is-**	'to say, make'
Burji	**iy-**, **i-**	'say, do'	**iss-**	'do, make'
PHEC	***y**	'to say'	***ys-** → ***is-**	'to say, make'
Had	**y-**	'say, be'	**iss-**	'do'
Kam	**y-**	'say'	**y-isiis-**	'to make say'
Sid	**y-**	'say, judge'	-	-
Ged	**hiyy-**	'say, try, do'	-	-

Several regular patterns emerge from the data. The Burji root is transcribed as *iy-* and *i-* with the causative *iss-*, with phonetic epenthesis of **-i-** in the stem, while the NHEC languages consistently show *y-* as the stem reflex. Gedeo *hiyy-* has inserted initial **hi**; therefore, Eyob Kelemework (2015, pp. 206, 82) observes that, in support verb constructions following nominal stems, the initial **hi-** is deleted (*hiyy-* → *-yy-*).[195] Kambaata **y-isiis-** 'make say' represents a morphological analysis of root + double causative (§3.1.2.3), contrasting with the causative forms **iss-** found in Burji and Hadiyya. This reflects an innovation in Kambaata and fusion in Burji and Hadiyya (**y + s** → geminate **ss**). Sidaama and Gedeo lack attested independent causative forms of **y*, though its use as the support verb persists (§4.2.4.3).

[195] A similar insertion process is found in Sidaama *hiyy-essa* 'wolf, jackal' from PHEC **yeyy-* 'wolf, jackal' (see §2.1.5.3 for initial **h** insertion).

The intransitive patterning of **y* has structural consequences throughout the denominative system. In §4.2.5.2, **y* serves as the foundation for denominal verbs with intransitive or middle semantics (inchoative, stative). The reflex **i-** appearing in causative denominatives like Sidaama *araar-(i)-s-* 'mediate, reconcile' (§4.2.5.7) preserves the support verb as a verbalizer ***y**, contrasting with the causative **-is** suffix on non-derived verbs where **-i-** is epenthetic (§3.1.2.2). This distinction enables denominative origin to be detected even when semantic compositionality has been obscured through lexicalization (§4.2.5.9 discusses semantic specialization).

The generalization pathway documented in §§4.2.5.6-4.2.5.84.2.5.7 argues that denominative constructions built on ***y** + causative **-s** (***y+s**) evolved into the synchronic causative morpheme **-is** through functional expansion, explaining why the causative **-is** is the most prominent, and it formally resembles the denominative pattern (i.e., due to their shared origin and convergence).

The HEC support verb **y* 'to say, do, be' has cognates throughout East Cushitic. Sasse (1982, p. 102) reconstructs PEC **-iy(y)-* 'say' based on widespread reflexes: Burji *iy-*, Sidaama *i-*, *iy-*, Gedeo, Kambaata, Hadiyya, *i-* 'say', Saho-Afar *y-* (suffixing), Afar *-iy(y)-* (prefixing), and possibly Dullay *iy-anna* 'inquit, that is, I mean'. Evidence for this verb persists in fossilized forms across Eastern Cushitic, such as in the verb [21] meaning 'to listen, hear'. The use of **y* as a verbalizer has also been reported for Southern Cushitic (Ehret, 1987, p. 9).

4.2.4.2 The Verb *ʔ 'to do'

The independent PHEC verb [425] **ʔ, *ass-* (caus) 'to do' is directly attested in Hadiyya and Kambaata. Burji, Sidaama, and Gedeo do not preserve **ʔ* as an independent lexical verb, though fossilized traces appear in derivational morphology (§4.2.5.1). Additional HEC cognates strengthen the reconstruction: K'abeena (Crass, 2005, p. 226), Alaaba (Schneider-Blum, 2007, pp. 277–278; 304–305), and Libido (Crass, n.d., p. 30) all attest reflexes of **ʔ* 'to do' (as cited in Treis, 2024, p. 165).

The following set presents the PHEC and PNHEC reconstructions for the verb **ʔ* 'to do' alongside attested reflexes in Hadiyya and Kambaata. The comparative evidence suggests that **ʔ* was the earliest form, with vocalic support developing independently in the daughter languages to satisfy phonotactic needs.

	Stem		**Causative**	
PHEC	***ʔ**	‘to do’	***ʔ-s-** → ***ass-**	‘to do’
PNHEC	***ʔ**	‘to do’	***ʔ-s-** → ***ass-**	‘to do’
Had	**aaʔ-**	‘do’	-	-
Kam	**aʔ-**	‘do’	**ass-, ees-**	‘do’
Sid	**-**	-	**ass-**	‘do’
Ged	**-**	-	**(h)ass-**	‘do’
Alaaba	**aʔ**	‘do’	**ass-**	‘do’
K’abeena	**aʔ-**	‘do’	**ass-**	‘do’
Libido	**aaʔ-**	‘receive’[196]	**-**	-

Allomorphic forms (*ʔ, aʔ, ʔa, a*) reflect variation in the position and realization of the glottal stop relative to the supporting vowel. The causative variants show **ass-*, derived from **ʔ-s-* via fusion and gemination with the causative ***s** (see §3.1.1). Gedeo *hass-* ‘do’ (parallel to *hiyy-* ‘say’) has initial **h** insertion. The distribution gap in Burji, Sidaama, and Gedeo reflects lexical replacement rather than morphological loss, as denominative uses persist (§4.2.5.1).

The restriction of **ʔ* to transitive contexts (contrasting with intransitive **y*, §4.2.4.1) is semantically coherent: ‘do’ inherently requires a patient argument, while ‘say’ and ‘be’ pattern intransitively in HEC. This transitivity distinction carries over to support verb contexts (§4.2.4.3) and denominative derivations (§4.2.5.1), where **ʔ* licenses transitive morphology.

4.2.4.3 Stem + Support Verb

As mentioned above, the two verbs can be used as support verbs when they occur in compounds following other stems (or attested nominal stems). They are found separately or directly attached to the preceding stem, where some morpheme boundary alternation processes may occur.

[196] Concerning the meaning of Libido *aaʔ-* ‘accept, receive’, Treis (2024, p. 165, fn. 24) rejects Crass’s (n.d., p. 30) claim that such a support verb used in transitive contexts in Libido goes back to ‘accept, receive’ in light of the comparative data and proposes that it might historically originate from the verb ‘to do’. Of course, Libido *aaʔ-* ‘accept, receive’ could be considered as a semantic extension, as evident in the derived PNHEC verb [32] **ʔ-ʔ-* ‘to take, receive’ → Hadiyya *haʔ-*, *aaʔ-* ‘receive, take, accept, buy’, Kambaata *aak’k’-* ‘receive, take, accept’, Sidaama *haad’-*, *aad’-* ‘take, receive’, and Gedeo *ad’d’-* ‘receive’, with optional initial **h** epenthesis and suffixation of one of the reflexive suffixes (§4.2.3). The stem of the reconstructed form **ʔ* resembles that of the verb **ʔ* ‘to do’ as do its reflexes. Most likely, it has received the reflexive ***-ʔ** in the derivation process, as evidenced by the double vowels and the fused consonant of the attached reflexive suffixes.

The lack of Burji, Sidaama, and Gedeo cognates in the previous section could be amended by provisionally inferring variants which occur as supporting verbs from some verb derivations, at least in Burji and possibly Sidaama, e.g., in Burji:

> *d'agal-* 'to return' → *d'agal-ʔ-* 'to make one's wife return after having left her',
>
> *mur-* 'to cut, harvest' → *mur-ʔ-* 'to delimit, define', *mur-am-* (pass) 'to be cut, harvested', *anga murʔ-* (phrase) 'to kill' (*anga* 'hand'),
>
> *koraa* (m) 'saddle' → *kor-ʔ-* 'to saddle',
>
> *daaba* (f) 'horizon' → *daaban-ʔ-* 'to go horizontally', *daabanee* (adj) 'horizontal', *daaban-s-* (caus) 'to lay horizontally, down' (Roba Dame & Wedekind, 2008).

The Burji reflexes in the examples prompt the inclusion of the variant *ʔ(a)-* as one of the realizations of PHEC **ʔ* 'to do' above.[197] Sidaama has even more tentative evidence for a potential parallel supporting verb *aa-* 'to make, take, give', e.g.:

> *ʃeemaate* 'agreement, pact' → *ʃeemaate aa* 'to make/conduct an agreement with a contractor', *ʃeemaata* 'to make an agreement/contract',
>
> *k'oricca* 'medicine, measure or solution' → *k'oricca aa* 'to take measure, to punish',
>
> *biddiʃʃa* 'guide, directive, department' → *bidiʃʃa aa* 'to give directives, instruction' (Indrias Xa'miso et al., 2007).

The dictionary definition of Sidaama *aa* is 'to give' only (Indrias Xa'miso et al., 2007, p. 60), but the meaning provided in the above examples suggests extending the definition to include 'to make' and 'to take'.

---

[197] The identical form of the reflexive ***-ʔ** and the verb **ʔ* 'to do' makes it difficult to distinguish them when the latter is used as a support verb. However, at least the semantics of some of the provided transitive verb examples do not fit the reflexive/middle semantics, so one can link them back to the verb **ʔ* 'to do'. The support verb also differs in the variants with a vowel, which are not typical of ***-ʔ**.

Semantically, the verb *y-* is intransitive while *aʔ-* is transitive in Kambaata (Treis, 2008, p. 87); *hiyy-* 'do' is intransitive while *ass-* 'say, do' is transitive in Gedeo (when they are part of derived/verbalized verbs) (Eyob Kelemework, 2015, p. 206); and *y-* 'to say' is intransitive while *ʔáss-* 'to do' is transitive in Sidaama (Yri & Pepper, 2019). They are summarized below.

	Intransitive		**Transitive**	
Burji	**iy-**	'say, be'	**iss-**	'do, make'
Kam	**y-**	'to say'	**aʔ-, ass-**	'to do'
Had	**y-**	'to say'	**aaʔ-, iss-**	'do'
Sid	**y-**	'say, try, do'	**ass-**	'do'
Ged	**hiyy-**	'to say, do'	**(h)ass-**	'do'

Accordingly, the root **y* 'to say, do, be' is inherently intransitive, while **ʔ* 'to do' is inherently transitive. The inherited semantics determine each verb's selection of valency suffix (e.g., **ʔ** + passive/causative and **y** + causative) when derived, so they have historical significance, which will be addressed in the following section (4.2.5).

Below is a summary of the two verb roots, their causative stems, and their common reflexes and realizations as independent and cliticized support verbs when attached to another stem:

Root		**Reflexes**	**Causative Stem**		**Reflexes**
***y**	'to say, do, be'	**y, i, iy**	***ys-** → ***is-**	'to make'	**iss-, is-**
***ʔ**	'to do'	**ʔ, a, aʔ, ʔa**	***ʔ-s-** → ***ass-**	'to do'	**ass**

A number of comparative sets exhibit such compounds (or verbalized stems) in the individual HEC languages. The examples show that, in compound forms with both **y* 'to say, do, be' and **ʔ* 'to do' functioning as verbalizers are found, sometimes alongside other forms, derived or non-derived, of the same stem. For instance, set [249] appears to be from **ʃik'-* 'to bring'.

[249]	***y 'to say, do, be'**	***ʔ 'to do'**	**Others**	
PNHEC	***ʃik' y**	***ʃik' ʔ**	**-**	'to approach'
Had	**-**	**ʃik'aaʔ-**	**-**	'approach, become close'
Kam	**ʃik' y**	**-**	**-**	'budge, move up'
Sid	**ʃik'k'i yi**	**-**	**ʃik'-, ʃik'k'iʃi**	'approach'
Ged	**ʃik'k' hiyy-, ʃikkiyya**	**ʃik'k'a**	**ʃik'-**	'approach, come close'

Where only one gloss is provided, it applies to all the stems, e.g., Sidaama and Gedeo. The Oromo parallel is *sik'a* 'move a little, drag on, push something'.

[31]	***y 'to say, do, be'**	***ʔ 'to do'**	**Others**	
PHEC	***oo y**	***oo ʔ**	**-**	'to cry'
Burji	**oy-**	**-**	**-**	'cry, weep'
PNHEC	***oo y**	***oo ʔ**	**-**	'to cry'
Had	**ooy-**	**-**	**-**	'to weep, mourn'
Kam	**-**	**oʔ-**	**-**	'cry, weep, mourn'
Sid	**ooy-, ʔoʔʔo yi**	**-**	**-**	'cry'
Ged		**-**	**ood'a, ood'-**	'wail, weep, mourn', 'cry'

Sasse (1982, p. 156) also reconstructs PHEC **oo y, *oo ʔ* 'to cry' as a compound PEC **ʔoo y* meaning 'say oo' (see Oromo *oo(y)-* 'cry').

[10]	***y 'to say, do, be'**	***ʔ 'to do'**	**Others**	
PHEC	***hat't'iʃ y**	**-**	**-**	'to sneeze'
Burji	**hattiʃeys-**	**-**	**hat't'iʃee-d'-, hat'iʃ-**	'sneeze', 'cause to sneeze', 'to sneeze'
PNHEC	***hat't'iʃ y**	**-**	**-**	'to sneeze'
Had	**hat'iʃʃuy-**	**-**	**hat't'eʃ-**	'sneeze'
Kam	**hit'iʃʃu y-**	**hant'iʃʃ-aʔ, hit'iʃʃo ass-**	**-**	'sneeze'

Some languages use either support verbs to verbalize the same stem, e.g., Kambaata *hit'iʃʃu y-*, *hit'iʃʃo ass-* (Hudson, 1989, p. 138) and *hant'iʃʃ-aʔ*, 'sneeze' (Treis, p.c.).[198]

[583]	***ʔ 'to do'**	**Others**	
PHEC	***maaz ʔ, *miiz ʔ**	**-**	'to hurt'
Burji	**-**	**miid'-**	'to harm, hurt'
PNHEC	***miiz ʔ**	**-**	'to hurt'
Had	**madaʔ-**	**-**	'to wound somebody'
Kam	**maz-ees- (← *maz ʔ-is)**	**maza-ta**	'wound somebody', 'wound' (n)
Sid	**-**	**miic'i**	'hurt'
Ged	**-**	**miid'a**	'hurt, wound somebody'

[198] Treis (p.c.) does not interpret this verb as an ideophone plus a light verb, but rather as a verb whose root ends in a fossilized/frozen middle marker **aʔ-**.

Burji, Sidaama, and Gedeo parallels in other stems are influenced by the Oromo parallel *miid'a* 'hurt, be spoiled', *manc'aʔa* 'hurt'.

[673]	***y 'to say, do, be'**	**Others**	
PHEC	***ik, *y k**	**-**	'to be(come), happen'
Burji	**ih-**	**-**	'to be; become'
PNHEC	***ik-, *ikk-**	**-**	'to be(come), happen'
Had	**ikkoʔo, ih-**	**-**	'to be, become, be enough'
Kam	**ih-**	**-**	'be, suffice'
Sid	**ik-, ikk-**	**-**	'become, be, happen'
Ged	**-**	**kada**	'become, be, happen'

The geminate variant in [673] PNHEC **ikk-* 'to be(come), happen' suggests that it is a combination of the verb **y* and the agent consonant *-k*, because the meanings of PHEC **y* include 'to be'. A similar combination can be seen in the Sidaama causative verb *see-kk-* 'to adapt, make suitable', which is, according to Yri & Pepper (2019), via an intermediate **see-y-k-* and derived from the stem *see-* 'to be suitable'.

Some stems occur in compounds with only one of the two support verbs, e.g., only **ʔ* occurs with PHEC [583]. It also remains unclear if there are semantic reasons other than transitivity for the use of either verb in compounds, e.g., Kambaata *c'aʔ(a) y-* 'be(come) clean/clear (of sky)' and *c'aʔ(a) aʔ-* 'make clear/clean (by removing everything)'.

Additional examples include: PHEC [170] **sum ʔ* → PNHEC **ʃum ʔ* 'urinate', [639] PHEC **ʃaf-*, **ʃaf-ʃaf ʔ* → PNHEC **ʃaf-*, *ʃaf-ʃaf ʔ* 'to shake', PHEC [214] **t'aban ʔ, *t'aban ʔ-s-* 'to slap' → PNHEC **t'aban ʔ-s-*, and PHEC [45] **fiik' ʔ* 'to whistle'.

It should also be mentioned that almost all sources have reported the use of the two verbs as support verbs, but only after ideophones. However, it is proposed in this study that, historically, their formal use has been more than simply deriving verbs from ideophones (§3.1.2.3, §4.2.1, §4.2.3).

The transitive patterning of **ʔ* determines its distribution in denominative derivation. In §4.2.5.1, **ʔ* serves as the basis for denominal verbs with transitive or causative semantics, enabling nominal concepts to participate in transitive argument structures. The reflex **a** that appears before various extensions provides diagnostic evidence for **ʔ* as the verbalizing element (see §4.2.5.9 reanalysis framework).

The grammaticalization pathway mostly centers on **ʔ* combinations:

> Etymology of passive **-am** (§4.2.1): ***ʔ** + **-m** → ***a-m** yields the passive denominal pattern. Section 4.2.5.6 proposes that the current passive **-am** has been generalized from an older complex denominative ***ʔ+m** in order to be used with (verbal) stems not derived from nominals, establishing the grammaticalization sequence: denominal construction → general valency morphology.
>
> Alternative causative **-is** (§4.2.5.7): ***y** + **-s** → ***-is** yielding causative denominatives. The frequent attestation of causative **-is** signals a generalized structure from an older complex denominative ***y+s** used with (verbal) stems not derived from nominals, confirming the grammaticalization sequence: denominal construction → general valency morphology. Sidaama shows optional **-i-** deletion (§4.2.5.7) reflecting reanalysis of the reflex of the support verb as phonological epenthesis, demonstrating ongoing grammaticalization.
>
> Agentive layering (§§4.2.5.3–4.2.5.4): The combinations ***y-t-**, ***ʔ-t-**, and ***ʔ-k-** establish denominatives with agentive semantics. The preservation of these consonants in reflexive fusion (***ʔ-k-ʔ, *ʔ-t-ʔ** → **-ak'k', -it't'**, §4.2.5.5) demonstrates a layered structure, where the extended support verb occurs between the stem and the reflexive, creating ternary morphological units that subsequently fuse.

Understanding the vocalic reflexes of **ʔ* and **y* as verbalizers (rather than epenthetic vowels) is a prerequisite for interpreting the unified system presented in §4.2.5.9 and how the verbalized nominal stems participate in standard verbal morphology.

The support verbs ***ʔ** and ***y** retain some lexical autonomy in the compound patterns above. The next section (4.2.5) traces the grammaticalization of ***ʔ** into a non-autonomous verbalizer that attaches to nominal stems, triggering layered derivational extensions (***-t, *-k,** reflexive ***-ʔ**) documented in §§4.2.5.1–4.2.5.9.

4.2.5 Denominative Derivation

The primary denominative derivational strategy involves the addition of a support verb (§4.2.4) to nominal stems, optionally followed by verbal suffixation, to derive verbs from nouns. The term "denominative" refers to morphological markers deriving verbs from nominal bases (see §1.5.2).

The support verb in denominative constructions often undergoes phonological reduction, leaving behind vocalic or consonantal traces. This section argues that the vowel elements **i**, **a**, and **ee** preceding consonantal derivation markers in HEC are not epenthetic but rather grammatical reflexes of the verbalizers **y* → **i** and **ʔ* → **a**, with **ee** arising through coalescence (**a** + epenthetic **-i-** → **ee**). This hypothesis accounts for otherwise unexplained vowel variation in the denominative morphology and connects synchronic patterns to the diachronic grammaticalization of support verbs into derivational affixes.

This section argues that synchronic valency-changing morphology in HEC (passive **-am**, causative **-is**) originated historically as complex denominative constructions that underwent grammaticalization through functional expansion. The formal similarity between denominative and valency-changing patterns reflects a shared origin rather than coincidence or analogy.

The analysis proceeds through progressive stages of grammaticalization:

> Foundation (§§4.2.5.1–4.2.5.2): Simple denominatives establish verbalizer patterns.
>
> Layering (§§4.2.5.3–4.2.5.4): Agentive suffixes add a second derivational stage.
>
> Complex Combinations (§§4.2.5.5–4.2.5.7): Reflexive, passive, and causative extensions build elaborate denominal derivations.
>
> Theoretical Synthesis (§4.2.5.9): The complete framework integrates reanalysis, two-stage derivation, generalization, and unification.

The cross-HEC variation reflects different stages and rates of grammaticalization: Kambaata preserves distinct agentive **-k**/**-h**; Gedeo and Hadiyya show fossilized and partially active agentive **-t**; Sidaama exhibits analogy-driven reanalysis of the verbalizer **-i** as epenthetic; Burji shows contact-influenced innovations alongside inherited patterns.

Each subsection below documents one stage of this pathway with explicit cross-references showing interconnections.

4.2.5.1 The Verbalizer *ʔ

The verbalizer ***ʔ*** (← support verb **ʔ* ‘to do’) is attested across all five HEC languages, though it appears in varying degrees of morphological

transparency. In Burji, ***ʔ** persists mainly in fossilized forms of verbalized stems where its verbalizer origin is no longer synchronically productive (i.e., patterns that cannot freely extend to new nominal stems (§4.2.4.2)). In contrast, Kambaata, Hadiyya, Sidaama, and Gedeo exhibit more transparent uses in which ***ʔ** combines productively with nominal stems to derive transitive denominal verbs (see below and §§4.2.4.2, 4.2.4.3 for more examples).

	Nominal [145]		**Verbalized** [170]	
PHEC	***sum-, *sim-**	‘urine’	***sum ʔ**	‘urinate’
Burji	**sinʔa, ʃinʔaa**	‘urine’	**sinʔ-aaw-, sinʔ-ay**	‘urinate’
PNHEC	***ʃum-**	‘urine’	***ʃum ʔ**	‘urinate’
Had	**ʃuma**	‘urine’	**ʃum-eʔ-**	‘urinate’
Kam	**ʃuma-ta**	‘urine’	**ʃum-aʔ-**	‘urinate’
Sid	**ʃuma**	‘urine’	**ʃumaa-ʔr-, ʃuma-ʔr-**	‘urinate’
Ged	**ʃiʔna**	‘urine’	**ʃiʔn-**	‘urinate’

In Sidaama, **ʔr** is a variant of **d’**, and the verbalizer is reflected as **aa** on the stem. There is a generalization of the verbalized stem to the nominal in Burji and Gedeo, hence the **ʔ** in their nominal stem. The Gedeo and Burji parallels are influenced by Oromo (*finc’-eessa*; *finc’-aay-* ‘urinate’ and *finc’aan* ‘urine’).

The reflexive and the verbalizer ***ʔ** are formally identical. However, the verbalizer ***ʔ** is traced back to a transitive verbal source (**ʔ* ‘to do’; see §4.2.4), in contrast to the middle/intransitive function of reflexive ***-ʔ**. Since it is originally independent of the preceding stem, the verbalizer ***ʔ** may be realized with a vowel, but the reflexive cannot. Practically, the reflexive suffix can be attached to the verbalizer, even though it will only surface as a single **ʔ** because there is no geminate **ʔʔ**.

This subsection establishes the first stage of two-stage denominative derivation: A nominal stem combines with the verbalizer ***ʔ** to create a verbalized base that can subsequently take standard verbal morphology (i.e., Stage 2: extensions **-t**, **-k**, **-ʔ**, **-m**, **-s**, as documented in §§4.2.5.3-4.2.5.7). The fundamental structure is:

> Stage 1: Nominal + Verbalizer ***ʔ** → Verbalized Stem
> Stage 2: Verbalized Stem + Extension → Complex Verbalized Stem

Typically, simple denominatives in ***ʔ** (without additional extensions) yield transitive verbs with causative semantics, reflecting the inherent transitivity of

the source support verb (§4.2.4.2). These patterns contrast systematically with the verbalizer **y**, which yields intransitive outcomes reflecting the intransitive nature of the original verb (§4.2.4.1).

The vowel **a-** appearing after stems and extensions is not epenthetic but grammatical, i.e., it is the reflex of verbalizing ***ʔ**. This reanalysis (detailed in §4.2.5.9) explains otherwise anomalous patterns: why the vowel quality is **a** rather than expected epenthetic **-i-** (§3.4.1.1), why the vowel persists even where phonotactics would not require epenthesis, and why verbal derivation suffixes like ***-s** show a preceding vowel consistently in the denominative.

The examples below demonstrate simple denominatives with the verbalizer **ʔ** across the HEC languages. For layered combinations (***ʔ** + extension), see §4.2.5.3 (agentive ***-t**), §4.2.5.4 (agentive ***-k**), §4.2.5.5 (reflexive), §4.2.5.6 (passive ***-m**), and §4.2.5.7 (causative ***-s**). For the generalization pathway showing how ***ʔ+m** evolved into synchronic passive morphology, see §4.2.5.6.

4.2.5.2 The Verbalizer *y

Treis (2008, p. 155) reports the use of the support verb *y-* 'say' as a verbalizer **y** in Kambaata in constructions involving ideophones. Comparable uses can be inferred for other HEC languages based on the data discussed below:[199]

	Nominal		**Verbalized**	
Burji	**monnokkee**	'dull'	**monnokku iy-**	'to be dull'
Had	**c'alk'a**	'handsome'	**c'alk'ay-**	'become handsome'
Kam	**wojj-u**	'white'	**wojj y-**	'(be)come white'
Sid	**taʃʃo**	'pleasure'	**taʃʃi y-**	'to be happy'
Ged	**ʃik'o**	'near'	**ʃikkiyya**	'be near'

Parallel to §4.2.5.1, this subsection documents the verbalizer ***y** built on the intransitive support verb **y* 'to say, do, be' (§4.2.4.1). The same two-stage derivational structure applies:

> Stage 1: Nominal + Verbalizer ***y** → Verbalized Stem
> Stage 2: Verbalized Stem + Extension → Complex Verbalized Stem

[199] Data sources: Kambaata (Treis, 2008, p. 155), Burji (Roba Dame & Wedekind, 2008, p. 63; see Tesfaye Baye, 2015, p. 198), Sidaama (Indrias Xa'miso et al., 2007, p. 553), Gedeo (Wedekind, 2008, p. 94), Hadiyya (Ritter, 2007, p. 24).

However, the inherent intransitivity of ***y** determines the semantic outcomes: denominatives in ***y** yield intransitive, inchoative, or stative verbs, contrasting systematically with the transitive/causative patterns of ***ʔ** denominatives (§4.2.5.1).

The key pattern of the verbalizer ***y** is the vowel reflex **i** appearing before stems and extensions, which reflects ***y**, not epenthesis. This is demonstrated by its consistent **i** quality across environments (if epenthetic, it would show contextual variation), its persistence even where not required by phonotactics, and its systematic contrast with **a-** (← ***ʔ**) in parallel environments.

The verbalizer ***y** receives the causative ***-s** to create complex patterns (§4.2.5.7) where the sequence ***y+s** yields the denominative causative **-is**. Section 4.2.5.8 demonstrates, through comparative data, that the denominative exponents adhere to a structural template (verbalizer + extension) retained from PHEC. The verbalizer ***y** also appears in agentive layering: ***y-t-** combinations (§4.2.5.3) yield denominals with agentive semantics. The fusion of agentive **-t** with reflexive **-ʔ** creates glottalized forms (***yt-ʔ** → **-it't'**, §4.2.5.5) where the intervening **t** is preserved, providing evidence for a ternary layered structure: verbalizer + agentive + reflexive.

According to the proposed grammaticalization, denominative patterns generalize (§4.2.5.9); they may extend to word classes beyond prototypical nominals, including sound-symbolic ideophones. The expansion of denominative morphology from nominal bases to other categories exemplifies the grammaticalization process documented throughout §4.2.5.

4.2.5.3 Agentive Denominative *-t

The archaic agentive verbal extensions **-k** and **-t** historically served to specify an active thematic role for the subject in denominative derivations. While the modern semantics of these verbs have largely generalized into inchoative ('become') or stative ('be') meanings, the presence of these consonants remains a fossilized indicator of an earlier stage where the verbalized stem explicitly encoded subject-agentivity.

Synchronic evidence for verbalized stems with the (agentive) consonant ***-t** is chiefly found in Gedeo. The Gedeo reflex is **-at** (Eyob Kelemework, 2015, p. 196; Dawit Tilahun, 2019, p. 206) as shown in the following examples:

	Nominal		Verbalized	
[256]	**hammeenna**	'rotten, bad'	**hamat**	'slander'
[1030]	**huk'k'aʃʃo**	'thin'	**huk'k'ata**	'be slim'
[128]	**d'eebo**	'thirst'	**d'eebota**	'to thirst'

The cognates in the other HEC languages have corresponding variants of the denominative suffixes, e.g., [128] Gedeo *d'eeb-ota* 'to thirst': Kambaata *t'eeb-ek-*, *t'eeb-eʔ-* 'be very thirsty', Hadiyya *t'eeb-eʔ-* 'to be thirsty'. The above examples have Oromo parallels: [256] *hammaat-*, *hamaa* (adj) 'bad off, unwell', [1030] *huk'k'at-* (adj) 'thin, skinny', and [128] *d'eebot-* (adj) 'thirsty'. The Oromo parallels listed in the dictionaries (Gragg, 1982; Tilahun Gamta, 2004) are nominal, mostly adjectives, but they share the identical suffix **-at**. Oromo also has variants of the same roots with the middle suffix **-d'(d')**.[200] The Gedeo and Sidaama parallels include Oromo loanwords, but it is unreasonable to rule out inheritance completely.

Sidaama stems ending with **-at**, e.g., [199], *ulat-* 'smell', [450], *saalfat-* 'be shy' (see *saal-* 'shame'), and [698] *konkolati* 'to roll', are rare and some of them might be borrowed from Oromo, e.g., *ulat-* (n), *ulannaa* (n) 'incensing or perfuming oneself with smoke of a plant that produces a sweet smell', *saalfateeraa* 'adolescence', and *konkolaataa* (n) 'car, vehicle (down hill)', respectively. Sidaama and Gedeo stems with **-at** rarely overlap (but see Gedeo *ulat-* 'smell good').

In Hadiyya, a sequence **-V-t** appears to be deducible in some cases, based on available data in Ritter (2007).

Nominal		Verbalized	
waaka	'greed'	**waak-at-**	'be greedy'
gaga	'self'	**gagi-itt-**	'be selfish'
hit'e	'meadow fescue grass'	**hit'-it-**	'to graze'

There has been a discussion in the early literature about Gedeo **-at**. Hudson (1976, pp. 272, 277) considers it to be another variant of the Gedeo reflexive and attempts to justify its existence alongside **-ad'-** by suggesting "[p]erhaps it is a back-formation based on the construction of certain nouns with adjectival meaning with the copula -te…, e.g., furda-te 'he is fat,' reanalysed

[200] Oromo **-t-** could be interpreted as a variant of middle voice verb extensions in Borana (Marsabit) and Wata dialects reported by Stroomer (1987, pp. 153, 155) which have variants **-ad'd'-** (for the first-person singular) and **-at-** (elsewhere). However, the latter seems to derive an adjective in *huk'k'at-* (adj) 'thin, skinny' in Oromo, while the Gedeo parallel is an inchoative *huk'k'ata* 'be slim'.

as a verbal construction, furd-at-e ‘he became fat,’ in which furd- is a basic stem, -at- a reflexive suffix, and -e the suffix of the 3rd sing.m. of the perfect”. Hayward (1984, pp. 89–90) rejects Hudson’s back-formation hypothesis. Given that the Oromo middle paradigm has a predominance of forms with **-t**, Hayward argues for an almost certain Oromo source for some of the examples cited by Hudson, supported by Hudson’s own note that Gedeo speakers are frequently bilingual, speaking Oromo and Gedeo. He also adds that “[f]rom the general standpoint of Highland Eastern Cushitic the forms with t are anomalous”.

Representing Stage 2 of the two-stage model, the agentive suffix ***-t** layers onto verbalized denominative stems created by the verbalizer ***ʔ** (§4.2.5.1) or ***y** (§4.2.5.2), yielding complex denominal verbs with agentive semantics.

> Nominal + Verbalizer (***ʔ/*y**) → Verbalized Stem
> Verbalized Stem + Agentive ***-t** → Complex Verbalized Verb

The provisional historical patterns would be: ***ʔ-t-** combinations yield transitive denominals with an agent/causer role, and ***y-t-** combinations yield denominals with an agent role but an intransitive frame.

The agentive ***-t** is not productively preserved in most of the HEC languages, but it appears in fossilized forms. Gedeo shows the clearest reflexes, while Hadiyya preserves traces in specific lexical items. Kambaata retained a distinct agentive ***-k** (§4.2.5.4), which potentially replaced or marginalized ***-t**. This distributional pattern suggests that ***-t** and ***-k** represent an earlier layer of PHEC morphology that has been largely superseded or lexicalized.

In addition to the evidence in Gedeo and Hadiyya, the preservation of ***-t** in reflexive combinations (§4.2.5.5) provides proof of layering. The fusion pattern ***y-t-ʔ** → **-it’t’** (agentive + reflexive → glottalized form) shows the agentive consonant intervening between the verbalizer and the reflexive. These ternary structures (verbalizer-agentive-reflexive) demonstrate that denominative derivation can involve multiple morphological layers that subsequently fuse into synchronic units. The preservation of **t** in reflexive contexts, even when lost elsewhere, exemplifies morphological fossilization.

There is provisional fossilized comparative evidence for the suffix ***-t** (§3.3.1.1) in the rest of NHEC, which complements the above **-t** in Gedeo and Hadiyya. However, a fossilized suffix **-s** is in competition with ***-t**, since they are both interpreted as complex denominatives that have lost the reflexes of the

preceding verbalizer ***y**, which would have otherwise prevented their fusion with the stem-final consonant.

[378]	**Fossilized -t**	**ø / -d'**	
PHEC	**-**	***dan-**	'to be able'
Burji	**-**	**dan-**	'be able'
PNHEC	***dan-t-**	**-**	'to be able'
Kam	**dand-**	**-**	'be able'
Sid	**dand-**	**-**	'be able'
Ged	**dand-**	**-**	'be able'

In set [378], the lack of final **d** in Burji compared to PNHEC **dan-t-* 'be able' might be due to the NHEC word having a fossilized denominative **(*y)-t** (assimilated in voicing to the final **n**), even though Oromo has a comparable form *dandaya* 'be able' (cf. Hadiyya *t'an-* 'be able').[201]

[29]	**Fossilized -t**	**b + t or s → f**	**ø / -d'**	
PHEC	**-**	**-**	***k'ab-**	'to hold, seize'
Burji	**-**	**k'af-**	**-**	'to have, possess, own'
PNHEC	**-**	**-**	***ab-**	'to hold, seize'
Ged	**abid-**	**-**	**-**	'hold, seize'

In this example, the verbalizer ***y** is preserved in Gedeo *abid-* 'hold, seize'; hence, **b** is not affected as in Burji.

As discussed in the chapter on the alternations at morpheme boundaries (§3.3.1.1), a suffixed ***(y)t** or ***(y)s** has conditioned the fricativization of root-final ***b** to **f**.[202] The regular intermediate would be geminate ***ff** from ***bs**, which is perceived as **f** because of the lack of voiceless **p** in these languages. Additionally, the single **f** reflex aligns with the lack of reconstructable geminate **ff**.

***b + *-t, *-s → *ff → f**

[201] See the following derived Hadiyya stem showing the array of derivational suffixes including **-tt**: *at'-itt-* (mid) 'to wear jewelry', *at'-it't'-am* (refl/pass) 'to beautify', *at'-is-* (tr) 'to make something good, to make something new, to repair', *at'-is-am-* (intr) 'to be correct', *at'-is-aan-co* (n) 'mender, repairer' (Ritter, 2007, pp. 12, 44)

[202] An alternative to the denominative (fossilized) **t** being the trigger for fricativization is the labial ***b** devoicing via the obstruent harmony hypothesis (fn. 36): ***b → f** when C_1 is a voiceless consonant.

For example, Kambaata has *huf-* ‘comb’ → *hup’p’-* (mid) ‘comb oneself’ and *t’oof-* ‘finish’, *t’oop’p’-* (mid) (Treis, 2022; Treis, p.c.).

4.2.5.4 Agentive Denominative *-k

The complex denominative suffix **-eek**/**-eeh** is attested only in Kambaata. Treis (2008) reports the form as **-eeh**, while Hudson (1989) transcribes it as **-eek**; the difference likely reflects analytical decisions rather than distinct morphemes.[203] The following set of data, mostly from Hudson (1989), presents Kambaata **-eek/-eeh**, along with the verbalizer variant **-e** + epenthetic **-ʔ-** where available in Hadiyya and Kambaata:[204]

	Nominal		**Denom -eek/h**	**Denom -ʔ**	
Kam	**t’eebo**	‘thirst’	**t’eebek-**	**t’eebeʔ-**	‘be very thirsty’
Had	**t’eebo**	‘thirst’	**-**	**t’eebeʔ-**	‘be thirsty’
Kam	**fooʃʃa**	‘breath’	**fooʃeek-**	**fooʃʃeeʔ-**	‘breathe’
Had	**fooʃʃa**	‘breath’	**-**	**fooʃʃeʔ-**	‘breathe’
Kam	**k’egu**	‘blood’	**k’egeek-, k’egeeh-**	**-**	‘bleed’
Had	**t’iiga**	‘blood’	**-**	**t’iigeʔ-**	‘bleed’
Kam	**maza-ta**	‘wound’	**mazeek-**		‘be wounded’
Had	**mada**	‘wound’	**-**	**madaʔ-**	‘to wound somebody’

Treis (2008, p. 272, fn. 23) notes in a brief footnote that “[t]he morpheme *-eeh* is a derivational morpheme which generates inchoative verbs on the basis of nouns; see, for instance, *gorr-ú* ‘hunger’ → *gorr-eeh-* ‘get hungry’”. Similarly, the adjective *haar-oo(-ta)* ‘new’ is related to the verb *haar-eeh-* ‘be(come) new’ but has no synchronic base form **haar-*.[205]

---

[203] According to Treis (p.c.), Hudson’s (1989) **-eek** is a misinterpretation because, in the perfect and perfective, **h** is strengthened and geminated to **kk**, which led Hudson to interpret this as a verb ending in **-k**. The verbal noun of these verbs clearly shows that the root ends in **h**, e.g., *mazeehu* ‘becoming wounded’. Within the context of this study, both analyses are acceptable because non-initial ***-k** becomes **-h** in all HEC, including Kambaata (§2.1.5.2 (PHEC *h)), while non-initial ***-h** has been lost throughout HEC (§3.2.2). Therefore, it is an analytical decision to represent the more abstract **-eek** or the more concrete **-eeh** – a useful disagreement for historical analysis.

[204] Data sources: Kambaata (Hudson, 1989, p. 82; Treis, 2008; Alemu Banta, 2016), Hadiyya (Hudson, 1989; Ritter, 2007).

[205] Treis (2008, p. 170) says, referring to examples such as the ones involving the inchoative, that “[t]he derivational formative by which a noun is created on the base of a certain verb is not predictable but conventionalized. So far hardly any conditioning factors for the choice of one or the other derivational pattern could be determined[…]. Nothing helps us to predict that nouns corresponding to the verbs [e.g., the de-nominal verbs *maz-eeh-* ‘be wounded’ ← *maza-ta* ‘wound’ and *gorr-eeh-* ‘be(come) hungry’ ← *gorr-u* ‘hunger’] share a stem with them (i.e. are “zero-derived”), are derived from them with a certain formative, or are themselves the base for the verb”. The answer to that

Further examples include:

> Kam *barr-eek-* ‘be late (of day)’, Burji *orr-aaw-* ‘be late (of day)’ (intr), *orr-aa-s-ad’-* ‘be late in morning (of person)’,
>
> Kam *gorru* ‘hunger’, *gorr-eeh-*, *gorr-ees-* ‘be hungry’

Hadiyya and Kambaata **e?** reflect the corresponding denominative, with a raised verbalizer vowel **-a** and an inserted glottal stop. The glottal stop is epenthetic after lenition (***-ek** → ***eh**) and the elision of an intermediate ***h** (see §3.2.2). Hadiyya shows a similar pattern of epenthesis, replacing an underlying consonant (see §3.3.4.1). In other words, an intermediate reflex **ah** of the verbalizer ***?** and the lenited agentive ***-k** yielded a compensatorily doubled (and optionally raised) vowel **ee**, resulting in **ee** and a recently suffixed **-k** – e.g., Hadiyya *mada?-* ‘to wound somebody’ (tr), verbalized with ***?**, not the agentive **-e-?-**, and Kambaata *mazeek-* ‘be wounded’ with agentive layer **-eek**. In this view, the variants with **e?** (underlying **eh**) in the above examples in Kambaata and Hadiyya reflect an intermediate stage before the addition of **-k**, e.g., Kambaata *t’eebe?-* ‘be very thirsty’, Hadiyya *t’eebe?-* ‘be thirsty’, and with support verb + **-k** in Kambaata *t’eebek-* ‘be very thirsty’.

An alternative hypothesis traces **-eek** to the coalescence of the verbalizer **a** with the verb [673] PHEC **ik-* → PNHEC **ik-*, **ikk-* ‘to be(come), happen’ (**a-** + **-ik** → **-eek**). However, no independent HEC evidence for **ik-* as a verbalizer component has been identified, and the simple morphophonological account (***-k** lenition + Kambaata compensatory doubling / Hadiyya epenthesis) requires fewer assumptions. The **ik-* hypothesis remains tentative pending further investigation.

In addition to the agentive suffix ***-t** (§4.2.5.3), the agentive ***-k** (realized as **-eek/-eeh/-e?** reflecting regular sound changes) represents Stage 2 of the two-stage model layering onto verbalized denominative stems created by the verbalizer ***?** (§4.2.5.1). It serves the same derivational function as ***-t** (§4.2.5.3) but remains productive in Kambaata. Its structure is parallel to ***-t** as follows:

---

puzzle, at least for the inchoative, would be that the nouns themselves are the base for the verb (see §4.2.3).

Nominal + Verbalizer (***ʔ**) → Verbalized Stem
Verbalized Stem + Agentive ***-k** (or [673] PHEC **ik-*) → Complex Verbalized Verb

Similar to Gedeo **-at**, **-eek/-eeh** appears robustly in Kambaata with inferable parallels in Hayiyya. This distribution pattern indicates functional replacement where **-k** occupies the derivational space that **-t** held in earlier stages. Alternatively, languages may have generalized one of the two, reducing the duality of the archaic morphology (assuming an earlier agent gender specification of the two).

Like ***-t**, the agentive ***-k** fuses with reflexive **-ʔ** to create a glottalized reflexive suffix via ***ʔ-k-ʔ** → **-ak'k'** in Kambaata and possibly Hadiyya (§4.2.3.2). This parallel fusion demonstrates that ***-k** follows the same structural template as ***-t**, occupying the middle position in ternary verbalizer-agentive-reflexive structures (§4.2.5.5). The replication of fusion patterns across different agentive suffixes suggests that the layered derivational template and the morphemes are inherited.

The distribution of agentive ***-t** (productive in Gedeo and marginal in Hadiyya) versus ***-k** (productive in Kambaata and inferable in Hadiyya) suggests functional replacement rather than phonological alternation: Both suffixes occupy the same derivational slot (post-verbalizer, pre-reflexive) and encode similar agentive/inchoative semantics, but show complementary language distribution. The shared Hadiyya and Kambaata fusion pattern (***-t-ʔ** / ***-k-ʔ** → glottalization, §§4.2.5.3–4.2.5.5)[206] suggests both suffixes are integrated into the same inherited template structure, favoring the functional replacement scenario over independent innovation. Subgrouping implications are examined in §5.3. Therefore, the agentive suffixes ***-t** and ***-k** are reconstructable to PNHEC, since Burji shows no reflexes.

The coexistence of scattered reflexes of ***-t** and ***-k** indicates that morphological replacement and erosion can occur at the level of individual derivational slots while preserving overall structural templates, productivity can shift between competing exponents of the same derivational function, and synchronic variation in NHEC reflects diachronic layering of ***-t** and ***-k**. This pattern of template preservation extends to the generalization process in

[206] Plazikowsky-Brauner (1957, p. 29) records a non-glottalized verbal reflexive suffix **-ak** for Hadiyya (§4.2.3.2).

§4.2.5.9, i.e., denominative templates (verbalizer + extension) generalize to valency-changing functions, while individual morphemes may be reanalyzed.

4.2.5.5 The Reflexive Denominatives

Several HEC grammars (Eyob Kelemework, 2015, p. 196; Treis, 2008, p. 271; Kawachi, 2007, p. 134; Tesfaye Baye, 2015, p. 197) report formal overlap between reflexive markers (§4.2.3) and the reflexive denominatives. The same morpheme (e.g., **ʔ**) appears on verbal stems and on nominal stems. This highlights the fuzzy formal and functional boundaries between markers of reflexives and denominatives in HEC. This section treats the nominal-stem cases as denominative, acknowledging their formal identity with the reflexives (§4.2.3), and proposing their historical trajectories.

	Form	**Nominal**		**Verbalized**	
Kam	**-ʔ**	**gaan-a**	‘fat’	**gaaʔn-**	‘be(come) fat’
		nubaacc-u	‘old’	**nubaap’p’-**	‘be(come) old’
	-it’t’	**gabbanc-u**	‘short’	**gabb-oo-t’t’-**	‘be(come) short’
Had	**-it’t’**	**hoffi**	‘small’	**hoff-ii-t’t’-**	‘become poor’
		gunda	‘short’	**gund-i-t’t’-**	‘shorten’
Ged	**-ʔ**	**haruma**	‘short’	**haruʔm-**	‘become short’
	-ed’	**k’eerra**	‘tall’	**k’eerr-ed’-**	‘become tall’
Sid	**-ʔ**	**duumo**	‘red’	**duuʔm-**	‘become red’
	-i-d’	**waajjo**	‘white’	**waajj-i-d’**	‘become white’
Burji	**-ʔ**	**gaani**	‘big’	**gaanʔ-**	‘become big’
	-ad’	**c’arbaá**	‘wet’	**c’arbaad’-**	‘become wet’
	-eed’	**duwwaaga**	‘red’	**duwweed’-**	‘become red’

The above examples show a strong resemblance between reflexive and denominative markers. However, they are analyzed as denominatives, and where they are identical to the reflexives, their denominative form is the older formation, which has been generalized to the reflexive.

Gedeo and Sidaama use several suffixes to derive intransitive denominal verbs from nominal stems. In Gedeo, these include **-ʔ**, **-at**, and **-ed’**; in Sidaama, **-ʔ** and **-i-d’**.[207] According to Eyob Kelemework (2015, p. 197), the

[207] Other forms or allomorphs are listed in the sources (e.g., Sidaama: **-ʔr**, **-p’**, **-t’** (Kawachi, 2007, p. 134)), but they are only products of phonological or morphophonemic processes.

distribution of these markers appears to be lexically determined rather than following a predictable phonological or semantic rule in Gedeo.[208]

Tesfaye Baye (2015, p. 183) specifies that **-eed'** with a double vowel is only used as a denominative in Burji. That vowel doubling is due to the coalescence of the two segments: a reflex of the verbalizer ***y** (§4.2.4.1) and the initial vowel of the reflexive suffix **-ad'** (§4.2.3.4). Additionally, Tesfaye Baye does not list the glottal stop as a denominative. However, there are a small number of examples with glottal stop **ʔ**, identified as middle forms in Roba Dame & Wedekind's (2008) Burji Dictionary. These are, however, better analyzed as verbalizers (even if they are underlyingly extended ***ʔ-ʔ**), such as *gaani* 'big' → *gaanʔ-* 'to become big, grow' with the causative **-s** replacing the glottal stop in *gaans-* (caus) 'to make big, enlarge'; similarly in *ʃaalla* 'shame' → *ʃaalʔ-* 'to be ashamed' compared with *ʃaallis-* (caus) 'to ashame'. The denominative is also formed by utilizing the verbalizer **y** in Burji *d'iid'aa* 'jealousy' → *d'iid'iy-* 'to become jealous' and before the causative (→ **i**) in *d'iid'is-* (caus) 'to make jealous' (Roba Dame & Wedekind, 2008).

Kambaata **-it't'** and **-ak'k'** are attested but not frequent as denominatives, and there is possibly even a rare form **-t** (§4.2.5.3), e.g., *gabban-c-u* 'short' → *gabb-oo-t't'-* (intr) 'be(come) short', *gabboot-is-* 'shorten' with **-i-** coalescence with a vowel-based reflex of the verbalizer ***ʔ** (§4.2.4.2). Due to their scarcity, they are not mentioned as independent suffixes in any of the previous literature. Likewise, Treis briefly mentioned the denominative **-eeh** (transcribed as **-eek** in Hudson (1989)) and **-ʔ** only as a middle marker obligatorily occurring on some verbs (Treis, 2008, p. 272).[209]

Hadiyya has further denominative suffixes discussed in the previous subsections (§§4.2.5.1–4.2.5.4); however, they are not documented in the literature and are mainly inferred from the available data.

The combination of verbalized stems with reflexive morphology reveals evidence for a layered morphological structure coming about by means of fusion processes. The standard reflexive patterns (baseline from §4.2.3) is the simple attachment of the basic reflexive ***-ʔ** (§4.2.3.1) to verbal stems with the

[208] Some Gedeo examples exhibit double derivation with the non-glottalized **-t** and the glottalized **-t't'** added to the adjective form: *hortaat't'e* (adj) 'fertile (of woman)', *hor-em-* (intr) 'breed, multiply', *hor-siis* 'breed' (tr) and *ildaat't'e* (adj) 'fertile (of woman)', *il-* (tr) 'give birth, beget'.

[209] According to Treis (2008, p. 272), crucially, several inchoative verbs derived from adjectives are "obligatorily middle-marked" in Kambaata.

expected regular alternation process in NHEC (§§3.3.1, 3.3.2). The complex denominative reflexives, on the other hand, include the support verb - agentive - reflexive ternary structures, and again, exhibit the NHEC alternation process affecting the intervening consonants.

The defining characteristic of denominative reflexives in Kambaata and Hadiyya is the preservation of agentive consonants (***-t/*-k**) in reflexive structures even when these consonants are lost or marginal elsewhere. The first pattern (***y-t-ʔ** → **-it't'**) is comprised of the verbalizer *y (§4.2.5.2) + agentive ***-t** (§4.2.5.3) + reflexive ***-ʔ** (§4.2.3.3, mostly Hadiyya). The agentive **-t** is fossilized in fusion, in addition to its preservation in non-reflexive contexts, in Gedeo and Hadiyya. The second pattern (***ʔ-k-ʔ** → **-ak'k'**) includes the verbalizer ***ʔ** (§4.2.5.1) + agentive ***-k** (§4.2.5.4) + reflexive ***-ʔ** (§4.2.3.3, mostly Kambaata). The agentive **-k** preserved in Kambaata and traceable in Hadiyya (§4.2.5.4) shows the same structural template as ***-t** in the reflexive.

The preservation of agentive consonants in fusion with the reflexive proves a ternary morphological structure. Denominative derivation involves three layers (support verb + agentive + reflexive), not just two (stem + reflexive). The agentive suffix slot is between the verbalizer and the reflexive. Morphological fossilization is evident in the preserved agentive ***-t** (§4.2.5.3), which is marginally inferred from the reflexive fusion. The same NHEC fusion pattern (verbalizer-agentive-reflexive → glottalized agentive consonant) applies to both ***-t** and ***-k**, proving that the derivational template is maintained regularly.

Non-denominative verbs take simple reflexive ***-ʔ** without intervening consonants: verbal stem + ***-ʔ** → reflexive verb (no suffix-internal complexity). The presence vs. absence of intervening agentive consonants thus serves as a diagnostic for the determination of denominative origin: Glottalized reflexives with **-ak'k'/-it't'** patterns betray underlying denominative structure even when synchronic speakers no longer analyze the forms as compositional (§4.2.5.9).

Over time, fused forms like **-it't'** undergo semantic specialization: The relationship between derived and base forms becomes less transparent. For example, a verb like Hadiyya *sigg-it't'-* 'to cool down' may develop idiosyncratic meanings beyond predictable "agent-reflexive" semantics. This semantic drift, coupled with formal fusion, obscures the underlying morphological complexity for synchronic speakers while preserving structural evidence for diachronic analysis.

The fusion processes documented here exemplify broader patterns in the grammaticalization pathway: Morphologically complex forms (ternary structures) undergo phonological fusion, fused forms develop specialized semantics diverging from compositional origins, original morphological boundaries become opaque to synchronic analysis, and structural remnants (preserved consonants) persist as diachronic traces. These same processes operate in the generalization of complex denominatives (***ʔ+m**, ***y+s**) into synchronic valency morphology (**-am** passive, **-is** causative): formal complexity → fusion → semantic specialization → categorial expansion.

The following is a summary of the complex denominatives/reflexives and the historical sources of their exponents discussed in this section:

Verbalizer	Agentive ***-t**	Agentive ***-k**	Reflexive **-(V)d'**
***ʔ**	**-at** (§4.2.5.3)	**-eek/-eeh** (§4.2.5.4)	**-eed'** (§4.2.3.4)
		↓	
	-	**-ak'k'** (§4.2.3.2)	**-ad'/-ed'** (§4.2.3.4)
***y**	**-it** (§4.2.5.3)	-	**-id'** (§4.2.3.4)
	↓		
	-i-t't' (§4.2.3.3)	-	-

The innovated reflexive **-(V)d'** is added as it is after the verbalizer in Burji, Gedeo, and Sidaama. The reflex of the verbalizer is more transparent in the double vowels of Burji **-eed'** and maybe Sidaama **-id'**. It should be noted that some of the instances that might seem to be simple verbalizers – i.e., they show the verbalizer **ʔ** only – might underlyingly be comprised of two glottal stops: the verbalizer **ʔ** plus the reflexive **-ʔ**.

4.2.5.6 The Passive Denominative

Gedeo and Burji show **-V(m)m** denominative forms extended with the passive ***-am**: Gedeo **-om** (Eyob Kelemework, 2015, p. 197) and Burji **-oom**, **-umm** (Tesfaye Baye, 2015, p. 198).[210]

[210] Data sources: Gedeo (Wedekind, 2008), Sidaama (Kawachi, 2007), and Burji (Tesfaye Baye, 2015; Roba Dame & Wedekind, 2008).

	Nominal		Verbalized	
Ged	**araarsaa**	‘peacemaker’	**araarema**	‘be reconciled’
	hiyyo	‘poverty’	**hiyyom**	‘become poor’
Sid	**araara**	‘reconciliation’	**araaram-**	‘become reconciled’
	k’arra	‘trouble’	**k’arram-**	‘be in trouble’
Burji	**haraaroo**	‘reconciliation’	**haraarmad’-**	‘become reconciled’
	hiyyuntaa	‘poverty’	**hiyyumm-**	‘become poor’
Had	**araassaanco**	‘mediator’	**araaram-**	‘make peace’ (intr)
	k’uunk’a	‘egg’	**k’unk’am**	‘be round’

There is a parallel **-am** in Sidaama even though Kawachi (2007, p. 335) does not recognize it as a denominative, instead noting that “[t]here are a few verbs that always contain the passive suffix and cannot be used as verbs without it, though they usually have related noun forms: k’arr-am- ‘to be in trouble’ (k’arra ‘problem, trouble’), araar-am- ‘to become reconciled’ (araara ‘reconciliation’, araar-s- ‘to reconcile’)”.

In Hadiyya, Tadesse Sibamo (2015, p. 145) only reports the use of **-aam** for deriving adjectives from nouns. However, the dictionary entries in Ritter (2007) include numerous denominal forms (marked as passive/reflexive) with **-am** matching the denominative in the other languages. Therefore, this study posits the denominative function for Hadiyya, as well.

The combination of denominative stems with the passive morpheme ***-m** establishes the etymological foundation for synchronic passive **-am**. The analysis here is crucial for understanding the generalization pathway detailed in §4.2.5.9.

The synchronic passive suffix **-am** across the HEC languages derives historically from a complex denominative construction: ***ʔ** + middle/passive ***-m** → ***a-m** → **-am**. At the first stage, the verbalizer ***ʔ** (§§4.2.5.1, 4.2.4.2) converts the nominal stem into a verbal base. At the second stage, middle/passive ***-m** (see §4.2.1 for basic passive patterns) attaches to the verbalized stem, yielding a denominalized passive with vowel **a** as a reflex of the verbalizer ***ʔ**, not as an epenthetic insertion. The complex denominative ***ʔ+m** originally derived passive verbs from nominal bases, enabling nominal concepts to participate in passive argument structures. The examples above illustrate this denominal pattern. The functions are compositional at this stage: denominal + passivization = passive denominative.

The comparative data in §4.2.1 suggest that the current passive **-am** is generalized from an older complex denominative ***ʔ+m** used with stems that

have not been derived from nominals. This establishes the grammaticalization sequence in two stages: original (a) nominal + ***ʔ+m** → denominal passive (restricted to verbalized stems), then generalized (b) any verbal stem + **-am** → passive (expanded to all transitive verbs). The generalization represents functional expansion, i.e., a derivational pattern originally restricted to denominative contexts extends to apply productively to any eligible verbal stem. The formal structure (**-am**) is preserved while the functional domain expands categorially.

Examples in §4.2.5.8 and §4.2.1 present comparative evidence for the diverse selections of the denominatives used on the same nominal base across HEC.

Understanding **a** as the reflex of the verbalizer ***ʔ** (rather than epenthetic) is essential for this interpretation. The reanalysis framework in §4.2.5.9 shows that what have traditionally been treated as epenthetic vowels in denominative constructions are in fact reflexes of these support verbs. Explaining the appearance of non-epenthetic vowels like **a** without recognizing ***ʔ** → **a** would obscure the proposed ***ʔ+m** → **-am**.

Burji and Gedeo show an alternative pattern: ***ʔ + w** (see the rounded vowels) → inchoative marking. This is attributed to Oromo contact (as a source for the **-w** based inchoative), thus representing local innovation via contact (Oromo influence on Burji/Gedeo) and a parallel structure: ***ʔ** + extension following the inherited template with subsequent concatenation of **-m**. Irregular rounding in the denominative construction due to the labial **-m** may alternatively explain the rounding.

The etymology of the passive morpheme **-am** exemplifies the central claim of §4.2.5, i.e., synchronic valency morphology originates in denominative constructions that generalize. The same patterns recur in causative **-is** (§4.2.5.7) and are synthesized in the theoretical framework presented in §§4.2.5.8–4.2.5.9 .

4.2.5.7 The Causative Denominative

A denominative derivational suffix -**is** is formally similar to the causative **-is**. Unlike the causative, where **-i-** is epenthetic, the denominative includes a reflex of the verbalizer ***y**, which surfaces as **i** or is deleted between consonants.

	Nominal		**Verbalized**	
Burji	**haraaroo**	‘reconciliation’	**haraar-s-**	‘mediate, reconcile’
Had	**araaro**	‘peace’	**araar-, araass-**	‘to reconcile’
Kam	**araaruta**	‘intervention’	**araar-s-**	‘mediate, intervene’
Sid	**araara**	‘reconciliation’	**araar-(i)-s-**	‘mediate, reconcile’
Ged	**araarsaa**	‘peacemaker’	**araar-s-**	‘mediate, reconcile’

Sidaama’s optional **i** in the verbalized *araar-(i)-s-* ‘mediate, reconcile’ reflects the verbalizer **y**, which has otherwise been lost throughout the HEC verbalized stems. Its elision in this environment may be motivated by analogy with the causative allomorph **-s** after single sonorants (cf. 3.1.2.2): Once **i** was reinterpreted as epenthetic, it could be deleted, since **-i-** is not obligatory between a stem-final single sonorant and **-s**.

The other verbalizer **ʔ** has also contributed to another common denominative **-ees(s)**, arguably through coalescence via ***ʔ-is** → ***a-is**. In this case, the causative variant **-is** (with epenthetic **-i-**) is preceded by the verbalizer **a-**.[211]

	Nominal		**Verbalized**	
Burji	**bayyinna**	‘distance’	**bayyeesa**	‘keep away’
Had	**eddecca**	‘clothes, clothing’	**edd-ees-**	‘to clothe somebody’
Kam	**kaakk**	‘cackling sound’	**kaakk-ees-**	‘cackle (of a hen)’
Sid	**wodan-**	‘heart’	**wodan-eess-**	‘make someone thoughtful’
Ged	**mada**	‘wound’	**madeess-**	‘wound’ (caus)

The geminated consonant **ss** in Sidaama and Gedeo is puzzling and might require a different analysis than the one posited here if they are the result of fusion of the causative **-s** with another consonant or extension with an additional causative (**-ees+-s**).

The following sets include two verbalized and extended forms of a lost source **ħuy-* ‘full’, using the complex (extended) forms of the denominative **ʔ-m** and **ʔ-s**:

[211] Data sources: Hadiyya (Ritter, 2007), Kambaata (Alemu Banta, 2016), Sidaama (Yri & Pepper, 2019), Gedeo (Eyob Kelemework, 2015).

	Denom ʔ-m- [520]		**Denom ʔ-s-** [607]	
PHEC	***ħuy** → ***ħuy ʔ-m-**	‘be(come) full’	***ħuy ʔ-s-**	‘to fill’
Burji	**huuʔ-, huum-**	‘be(come) full’	**huuss-, huus-**	‘fill’
PNHEC	***ħuy ʔ-m-**	‘be(come) full’	***ħuym-s-**	‘to fill’
Had	**woʔm-**	‘be(come) full’	**wanʃ-, wonʃ-**	‘fill’
Kam	**wiim-, yiim-**	‘be(come) full’	**wiinʃ-, yiinʃ-**	‘fill’
Sid	**woʔm-**	‘be(come) full’	**wonʃ-, woʔm-is-**	‘fill’
Ged	**woʔm-**	‘be(come) full’	**wonʃ-, wanʃ-**	‘be full, fill’

PHEC **ħuy-* ‘full’ becomes **ħuy ʔ-m-* ‘be(come) full’ and **ħuy ʔ-s-* ‘to fill’, while PNHEC **ħuym-s* ‘to fill’ uses the verbalized stem (instead of the root **ħuy-*) as a derivational base, upon which it adds the causative **-s** without a need for a verbalizer. Sidaama even generalized the verbalized *woʔm-* ‘be(come) full’, so it uses the causative with epenthesis (**-i-s**) in *woʔm-is-* ‘fill’ because of the stem-final cluster (see §4.2.2.2). Kambaata *wiim-*, *yiim-* ‘be(come) full’ and *fiil-* ‘unravel’ might have lost the fossilized glottal stop and compensated for that loss by vowel doubling. In addition to the above, see Burji *huuwʔa* ‘to be filled up, be full of’ (Getahun Addo, 2017, p. 478) and Gedeo *uwwa* ‘full’ (Dawit Tilahun, 2019, pp. 59, 55), with the assimilation of stem-final **y** to **w** after **u**.

The combination of denominative stems with the causative **-s** reveals multiple etymological pathways that converge on formally similar surface patterns. The analysis establishes critical foundations for the generalization framework in §§4.2.5.6, 4.2.5.7, and 4.2.5.9.

Two distinct patterns emerge in the causative denominative. First, **-is** with ***y** reflex (← ***y+s**) via nominal + ***y** (intransitive support verb) → verbalized stem, verbalized stem + causative ***-s** → denominative causative. This is the transparent denominative causative with the verbalizer preserved as **i**, distinguishing it from the simple causative **-is** on non-derived verbs (where **-i-** is epenthetic, see §4.2.2.2). Second, alternative causative **-ees** via coalescence (← ***ʔ+is** → ***a-is** → **-ees**) via nominal + ***ʔ** (← transitive support verb) → verbalized stem + causative **-is** → denominative causative. This pattern shows the verbalizer ***ʔ** (§4.2.5.1) combining with the causative suffix **-is**.

The denominative causative patterns documented here provide the etymological source for the synchronic causative morphology. Sections §4.2.2.2 and §4.2.5.7 suggest that the denominative/derived support verb **-is**

from ***y+s** has been generalized to the standard causative derivation. This means that the original function of the denominative causative (restricted to verbs derived from nouns) has been generalized to the general causative (which applies to any verbal stem). The formal **-is** structure is maintained through functional expansion, and the original verbalizer reflex (**i** ← ***y**) is sometimes dropped, enabling analogical processes like Sidaama deletion or reinterpretation as causative **-s** (§4.2.2.1).

The generalization pathway for causative **-is** parallels that of the passive **-am**: in the passive, ***ʔ+m** (denominative) → **-am** (general valency marker) and in the causative, ***y+s** (denominative) → **-is** (general valency marker). Both represent functional expansion from denominal derivation to general valency alternation, preserving formal structure (***ʔ** → **a,** ***y** → **i**) while expanding categorial distribution.

Synchronically, the two sources of surface **-is** in the HEC languages are the denominative **-is**, which contains verbalizer reflex ***y** → **i** (grammatical), and the causative **-i-s** with epenthetic **-i-** (phonological). The denominative cannot delete **-i-** without destroying derivational transparency; however, in the causative, **-i-** is subject to the epenthesis rules set in §3.1.2.2.

The deletion of the deverbal **y/i** in denominative causatives represents the collapse of this distinction. As **-is** generalizes, the denominative pattern is reanalyzed as phonological, enabling both patterns to converge formally and functionally.

The following summary presents the historical sources of the complex passive and causative denominatives:

Verbalizer	Passive ***-m**	Causative ***-s**
ʔ	**-am** (§4.2.1)	**-ees** (§4.2.2.2)
y	-	**-is** (§4.2.2.1)

The overall generalization pathway is synthesized in §4.2.5.9.

4.2.5.8 Comparative Data

This subsection presents comparative data demonstrating the complex denominative constructions discussed in the previous section.

The following examples illustrate the role of different denominative morphemes in deriving verbs from nominals, as documented in the

comparative database. The glosses are provided directly beneath each lexeme. The verbalizers in the first row of the table have different vocalic reflexes when followed by one of the extension consonants (**-t**, **-k**, **-s**, **-m**). Otherwise, they may surface as the consonants **ʔ** or **y** when not followed by any of those consonants, as illustrated in [392] and [583].

Example [426] below presents comparative data for the reconstruction of denominative derivations. Using PHEC [426] **ħaraar y-s-, *araar y-s-* as a diagnostic case, the table tracks the selection of verbalizing strategies across the languages.[212]

[426]		**Verbalizer ʔ**	**Verbalizer y**	**Nominal Stem**
Burji	**-s**	-	*haraar-s-* 'mediate, reconcile'	*haraaroo* 'reconciliation'
Had	**-s**	-	*araass-* 'to reconcile'	*araaro* 'peace'
Kam	**-s**	-	*araar-s-* 'mediate, intervene'	*araaru-ta* 'intervention'
Sid	**-s**	-	*araar-i-s-*, *araar-s-* 'to reconcile'	*araara* 'reconciliation'
Ged	**-s**	-	*araar-s-* 'mediate, reconcile'	*araarsaa* 'peacemaker'
Had	**-m**	*araar-am-* 'make peace' (intr)	-	-
Sid	**-m**	*araar-am-* 'to become reconciled'	-	-
Ged	**-m**	*araar-em-* 'be reconciled'	-	-

The data distinguishes between the causative denominative and passive denominative, e.g., in Hadiyya. While ***ʔ** yields regular **a**/**e**, the vowel reflexes of the verbalizer ***y** vary between the optional remnant **i** in Sidaama and loss in the rest of HEC, as discussed in §4.2.5.7.

[212] Data sources: All languages (Hudson, 1989), Sidaama (Kawachi, 2007), Hadiyya (Ritter, 2007), Gedeo (Wedekind, 2008; Gasparini, 1995, p. 95), Kambaata (Treis, 2008; Alemu Banta, 2016).

[392]		**Verbalizer ʔ**	**Verbalizer y**	**Nominal Stem**
Had	**ø**	-	**lik'aayy-** 'to borrow something'	**lik'aayyo** 'loan, debt, lending'
Burji	**-t**	**lik'eesad'** 'borrow'	**-**	**lik'ee** 'loan'
Had	**-t**	**-**	**lik'aayyit't'-** 'to borrow'	**-**
Sid	**-t/d'**	**-**	**lik'-iʔr-is-** 'to lend (sth. cannot be returned as it is)'	**-**
Ged	**-t**	**lik'eesseda** 'to borrow'	**-**	**-**
Burji	**-s**	**lik'ees-** 'lend (money)'	**-** -	**-**
Had	**-s**	**-**	**lik'aayyis-** 'lend'	**-**
Kam	**-s**	**leʔees-** 'lend of what cannot be returned'	**-**	**-**
Sid	**-s**	**-**	**lik'iiss-** 'lend'	**-**
Ged	**-s**	**lik'eess-** 'lend (money)'	**-**	**lik'a** 'debt, borrowed property'

The stem **lik'-* 'to lend' shows multiple derivational strategies across languages, including simple denominative in Hadiyya *lik'aayy-*, reflexive denominative in Burji *lik'eesad'-* and Hadiyya *lik'aayyit't'-*, and causative denominative in Sidaama *lik'iiss-* 'lend'.

[285]		**Verbalizer ʔ**	**Verbalizer y**	**Nominal Stem**
Burji	**-s**	**-**	**balʔ-is-** 'widen'	**balʔaneega** 'wide' (adj.m)
Ged	**-s**	**-**	**bad'd'-is-** 'widen'	**baʔla** 'wide, broad'
Burji	**-t/d'**	**balʔad'-** 'to become wide, broad'	**-**	**-**

[128]		Verbalizer ʔ	Nominal Stem
Had	**ø**	**t’eebeʔ-** ‘to be thirsty’	**t’eebo** ‘thirst’
Kam	**ø**	**t’eebeʔ-** ‘be very thirsty’	**-**
Burji	**-t/d’**	**d’aybood’-** ‘to become thirsty’	**d’ayboo** ‘thirst’
Ged	**-t**	**d’eebota** ‘thirst’	**deebo** ‘thirst’
Kam	**-k**	**t’eebek-** ‘be very thirsty’	**-**

[583]		Verbalizer ʔ	Verbalizer y	Nominal Stem
Had	**ø**	**madaʔ-** ‘to wound somebody’	**-**	**mada** ‘wound’
Sid	**ø**	**miic’i** ‘hurt’	**-**	**-**
Ged	**ø**	**miid’a, madoʔ-** ‘hurt, wound somebody’, ‘wound’	**-**	**mada, madoʔa** ‘wound’
Kam	**-k**	**maz-eek-, maz-eeh-** ‘be wounded’	**-**	**maza-ta** ‘wound’
Kam	**-s**	**maz-ees-** ‘wound somebody’	**-**	**-**
Sid		**-**	**mad-iiss-** ‘to wound’	**-**
Ged	**-s**	**madeess-** ‘wound’	**-**	**-**

The comparative data above demonstrate that denominative derivation in the HEC languages involves a range of suffixes, including causative, passive, and reflexive markers. The distribution of these suffixes and the reflexes of the verbalizers highlight the complex historical layering. The evidence from Burji, Gedeo, Sidaama, and Kambaata supports the view that denominal derivation is both morphologically diverse and sensitive to language-specific changes. Further research is needed to refine the criteria for distinguishing the function of each denominative in HEC verb formation.

4.2.5.9 Summary and Remarks

This section synthesizes the denominative analysis presented in §§4.2.5.1–4.2.5.8, articulating the complete theoretical framework that accounts for the

formal, arguably functional, and distributional properties of HEC denominative and valency-changing morphology. They are summarized in five key insights.

Reanalysis Framework

The analysis establishes that vowels appearing before verbal extensions are not phonologically epenthetic but grammatically meaningful, serving as reflexes of verbalizers: **i** ← ***y** (from the intransitive support verb, §§4.2.4.1, 4.2.5.2), **a** ← ***ʔ** (from transitive support verb, §§4.2.4.2, 4.2.5.1), and **ee** ← coalescence of ***ʔ** + **i** → **-ees** (§4.2.5.7) or extended ***y**.

The denominative system displays a clear structural pattern in which support verbs function as core elements that enable category change. This pattern supports the analysis that what have traditionally been treated as epenthetic vowels in denominative constructions are in fact reflexes of these support verbs, explaining the appearance of non-epenthetic vowels like **a** and **ee**.

The epenthesis analyses predict that vowels are inserted to satisfy phonotactic constraints (§§3.4.1, 3.1.2.2, 3.1.1.5). They should show contextually predictable quality (e.g., high vowel **-i-** in most HEC languages). They should appear only where phonotactics require them (e.g., for breaking illegal consonant clusters). They should exhibit regular reflexes in quality. However, denominative vowels show non-predictable quality, with **a** appearing where epenthesis would predict **-i-**; non-required contexts, with vowels persisting even where phonotactics allow consonant sequences; and systematic distribution, with **i** correlating to ***y** for cases of intransitive semantics and **a** correlating to ***ʔ** for cases of transitive semantics. This distribution is explainable if the vowels are grammatical (verbalizer reflexes) rather than phonological (epenthesis).

Treating **i/a** as grammatical enables the reconstruction of denominative etymology even in synchronically opaque forms. For example, passive **-am** transparently derives from ***ʔ+m** once **a** is recognized as grammatical (from ***ʔ**), whereas its analysis as epenthetic would obscure this etymology.

Two-Stage Derivation Model

Denominative formation in PHEC involved a two-stage process:

> Stage 1 – Verbalization: Nominal Stem + Verbalizer (***y** or ***ʔ**) → Verbalized Stem

The verbalizer enables category change from noun to verb, and the transitivity of the source support verb determines the argument structure of the derived verb, i.e., broadly ***y** with intransitive/inchoative meanings and ***ʔ** with transitive/causative meanings.

> Stage 2 – Extension: Verbalized Stem + Verbal Morphemes **(*-t, *-k, *-ʔ, *-m, *-s**) → Complex Deverbalized Stem

The standard verbal extensions that attach to the verbalized stems are agentive (§§4.2.5.3–4.2.5.4), reflexive (§4.2.5.5), passive (§4.2.5.6), and causative (§4.2.5.7). The result is ternary structures like ***y-t-ʔ** (verbalizer + agentive + reflexive) → synchronic **-it't'** via fusion.

The preservation of intervening consonants in reflexive fusion (§4.2.5.5) proves layering: forms like **-it't'** (§4.2.3.3) and **-ak'k'** (§4.2.3.2) show agentive ***-t** and ***-k** between a support verb (***y** or ***ʔ**) and the reflexive morpheme (***-ʔ**), demonstrating a three-layer structure rather than a binary stem + reflexive.

The two-stage model explains why denominative extensions appear after vowels (Stage 1 output) rather than directly on nominal stems and why denominal verbs can layer multiple meanings (e.g., agentive + reflexive → agent-reflexive). Languages preserve/lose/adapt at different stages (Kambaata preserves Stage 2 transparency via **-k**; Hadiyya fossilizes Stage 2 via **-t**). Stage 2 extensions (originally specific to denominal bases) generalize to apply to any verbal base.

Generalization Mechanism

The central theoretical claim is that synchronic valency-changing morphology originated as complex denominative constructions that generalized. The original function (denominal) was restricted to verbs derived from nominal bases: ***ʔ+m** → passive denominative (verbalized → passive verb) and ***y+s** → causative denominative (verbalized → causative verb). Then, the generalized function (valency) expanded to apply to any eligible verbal stem: **-am** → general passive (any transitive verb → passive) and **-is** → general causative (any root → causative).

The formal similarity between denominative markers and other verbal morphemes is due to the cliticization of the complex denominative to the nominal stem and to the generalization of some of those denominatives into

the reflexive, causative, and passive forms (e.g., **-am** ← ***ʔ-m**) due to functional similarity. The reverse pathway of verbal suffixes generalizing to nominal domains is also attested (e.g., glottalized reflexives used as denominatives).

The passive denominative (***ʔ+m** on verbalized stems) and general passives (valency alternation on any transitive verb) share a valency-reducing function. This functional overlap enabled categorial expansion; for example, once speakers recognized functional equivalence, the formal pattern extended beyond the original denominal restriction.

Therefore, numerous valency morphemes are better analyzed as generalized denominative patterns. The formal structure (verbalizer + extension) is preserved through functional expansion.

Unified System

The preceding analysis supports a unified system: support verbs grammaticalize as verbalizers, enabling nominal stems to participate in standard verbal morphology.

The denominative system and valency morphology form a continuum rather than discrete categories. The most transparent is the simple denominatives (***ʔ**, ***y** alone, §§4.2.5.1–4.2.5.2) with compositional noun → verb meaning. The intermediate is the complex denominatives (***ʔ**+extension, ***y**+extension, §4.2.5.3–4.2.5.4) with a layered structure. The most grammaticalized are the general valency markers (§§4.2.5.5–4.2.5.7) with an expanded functional domain (§§4.2.1–4.2.3).

All points on this continuum share the fundamental structure: Verbalizers enable category change from noun → verb (denominal) or are utilized within the verbal system (valency alternation).

This unified architecture contrasts with analyses treating denominatives and valency morphology as unrelated subsystems. The historical connection motivates synchronic integration.

Semantic Specialization and Lexicalization

The semantics of the denominalized stems is not always transparent. The causative denominative often adds semantic components beyond simple causation. This is most evident in cases like example [392] in §4.2.5.8, where

the relationship between derived and base forms is less predictable than in standard causatives.

Older lexicalized denominals show semantic drift. The meanings diverge from compositional predictions. Newer generalized valency uses show compositional semantics, e.g., the passive **-am** and causative **-is** add predictable meaning to verbal bases. Semantic opacity is diagnostic of historical depth, as non-transparent denominals represent earlier formations that have lexicalized, while transparent valency uses represent more recent generalizations, maintaining compositionality.

Therefore, not all formal patterns with verbalizer reflexes (**i**, **a**) have transparent denominal meanings synchronically. Morphemes (***y**, ***ʔ**) can be reconstructed from formal structure even when semantic compositionality has been lost.

The cross-linguistic variation in denominative marking within HEC reflects different historical trajectories following language divergence from PHEC. Kambaata preserves more distinct forms of the agentive **-k**/**-h** (§4.2.5.4), showing retention of ***-k** as a productive agentive, transparency in Stage 2 layering, and resistance to generalization/lenition in some patterns. Gedeo and Hadiyya show evidence of the other agentive ***-t** (§4.2.5.3), indicating the retention of a PHEC suffix. Sidaama exhibits transparent optional retention of the verbalizer **i** in a position where the other languages show analogy-driven reanalysis (§4.2.5.7) by dropping it, treating the denominative reflex as phonological (epenthetic). This collapses the denominative/valency formal distinction. Burji, Gedeo, and Sidaama show contact-influenced innovations in the reflexive denominative (§4.2.5.5), attaching the borrowed **-(V)d'** to the verbalized stems as an innovative derivation alongside inherited patterns. A separate branch status (not PNHEC) is visible in differential denominative patterns, e.g., Kambaata and Hadiyya patterns in §§4.2.5.7, 4.2.3.3.

In summary, each language preserves different stages and aspects of the original PHEC denominative system, enabling reconstruction through triangulation: Kambaata evidence for ***-k**, Gedeo/Hadiyya evidence for ***-t**, universal evidence for ***ʔ/*y** verbalizers, and differential generalization patterns documenting the timeline of grammaticalization.

The following two schemas present the reconstructed pattern and the surface reflexes attested in the HEC languages, with cross-references to subsections showing where each pattern is documented:

Verbalizer	Agentive ***-t**	Agentive ***-k**	Refl **-ʔ**	Pass ***-m**	Caus ***-s**
ʔ	**-at** (§4.2.5.3)	**-eek/-eeh** (§4.2.5.4)	**-ʔ** (§4.2.3.1)	**-am** (§4.2.1)	**-ees** (§4.2.2.2)
	-	-	**-ak'k'** (§4.2.3.2)	**-**	**-**
y	**-it** (§4.2.5.3)	-	-	-	**-is** (§4.2.2.1)
	-	-	**-i-t't'** (§4.2.3.3)	-	

The reconstructed combinations show the original verbalizer + extension sequences.

Verbalizer	Agentive ***-t**	Agentive ***-k**	Refl ***-ʔ**	Pass ***-m**	Caus ***-s**
ʔ	***ʔ-t-**	***ʔ-k-**	***ʔ-ʔ-**	***ʔ-m-**	***ʔ-i-s-/*ʔ-s-**
	-	**-**	***ʔ-k-ʔ-**	**-**	**-**
y	***y-t-**	**-**	**-**	**-**	***y-s-**
	-	**-**	***y-t-ʔ-**	**-**	**-**

The two schemas present the full combinatorial space of verbalizer (***ʔ, *y**) with extensions (***-t, *-k, *-ʔ, *-m, *-s**), showing which combinations are attested (with cross-references), which combinations are unattested (marked with "-"), how surface forms relate to proto-reconstructions, and where detailed discussions appear in the subsection of §4.2.5. The systematic gaps (unattested combinations) constrain the reconstructed system, showing that not all logically possible combinations existed. Attested patterns show that layering follows principled constraints, not free combination. **Figure *4*** summarizes the verbal derivation morphemes (§4.2) and their development, highlighting the role of the denominative constructions.

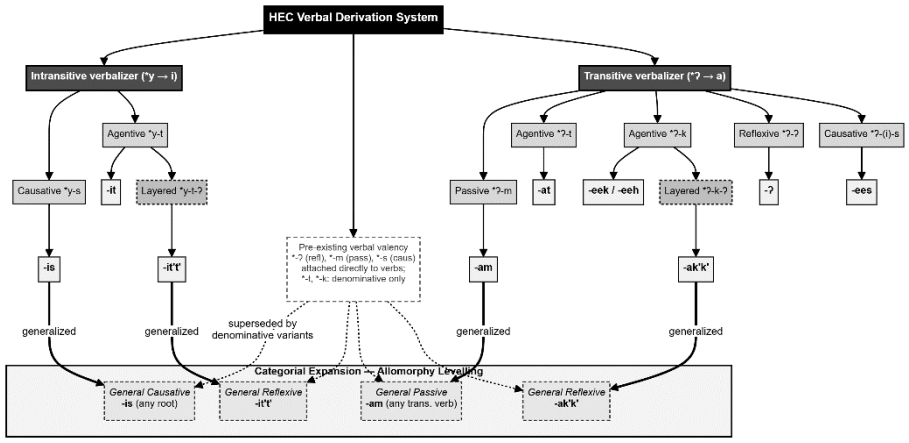


Figure 4: *The Development of the HEC Verbal Derivation System*

The preceding subsections identify the principal verbal derivational markers that can be compared across HEC and evaluate which forms are reconstructable. The comparative distributions show both continuity in the core derivational inventory and language-specific changes in particular markers and combinations. Where the evidence is uneven or restricted, reconstruction remains provisional.

The next section synthesizes the nominal and verbal reconstructions and presents a comparative summary of the morphemes discussed in this chapter. The goal is to make the distribution of reconstructed markers explicit and to highlight the most diagnostic morphological evidence for the classification developed in Chapter 5.

4.3 Summary and Conclusions

This section synthesizes the nominal and verbal reconstructions developed in this chapter and highlights the morphological evidence that bears on the subgrouping argument presented in Chapter 5. **Table 7** provides a comparative index of the nominal and verbal morphemes discussed in this chapter, and **Table 8** summarizes the reconstructed personal pronoun paradigm. The distributions support two generalizations: (a) much of the core derivational architecture is reconstructable to PHEC even when individual exponents show conditioned allomorphy or uneven attestation; and (b) a smaller set of shared innovations within NHEC provides diagnostic morphology for the Burji vs. NHEC split and for internal NHEC relationships.

Where the evidence is restricted or ambiguous, the proto-level assignment (PHEC vs. PNHEC) remains provisional.

Table 7 is intended as a comparative index to provide a compact basis for evaluating proposed PHEC reconstructions against PNHEC innovations by making comparative evidence and distribution explicit. Conventions are as follows: An asterisk marks reconstructed forms; hyphens indicate bound morphemes; and a dash indicates that a morpheme is unattested or not recoverable for the relevant language on the basis of the evidence assembled in this chapter. Where a marker shows conditioned allomorphy, the table lists the most general exponent and gives major alternants in parentheses; conditioning is discussed in the relevant subsection.

Category	PHEC	Burji	PNHEC	Hadiyya	Kambaata	Sidaama	Gedeo
Nominal							
Singulative	***-ʃ (*ʃ-ʃ, *-i-ʃʃ, *-c)**	-iccoo, -coo, -ʃo(o), -ʃaa	***-ʃ (*ʃ-ʃ, *-i-ʃʃ, *-c)**	-icco, -co	-iccu, -cu, -eeccu, ʃʃu/ccu	-icco, -co, -ʃʃo	-icco, -co, -ʃo, -jo
Reduplication plv	***C_i~C_i-a**	-	***C_i~C_i-a**	-a	C_i~C_i-a	C_i~C_i-a	C_i~C_i-a
Plurative	***-uw-a**	-	***-uw-a**	-uwwa	-	-uwa	-uwwa
Collective	***-ey-e**	-aayyaa, -aayyee	***-ey-e**	-	-	-e(e)yye	-eeye
Plurative	***-n**	-na	-	-	-	-aano	-
Deverbative	***-ʃ**	**-aancoo**, **-issoo**	***-ʃ**	-anca, -ca, -i-ʃʃa	-c-a(-ta), -iʃʃ-a(-ta)	-aan-co, -an-ʃo	-an-jo, -acco
Genitive	***ʔ**	-ng, -nd, -kk, -tt	***-ʔ**	-ʔ, -C'C'	-V́(V) (excluding u)	-ú, -í, -te, -◌́, -i, -ø	-nk, -k, -nt', -(t')t'
Masculine	***k**	k	***k**	k	k	k	k
Feminine	***t**	t	***t**	t	t	t	t
Masc. (nom)	***-u ← *-w**	-u (def)	***-u**	-i	-u (dem)	-u, -i	-i
Fem. (nom)	***-i ← *-y**	-i (def)	***-i**	-i	-i (dem)	-	-
Masc. (acc)	***-o ← *-óo**	-	***-o**	-o (TV)	-a (TV)	-o (TV)	-a (TV)
Fem. (acc)	***-e ← *-ée**	-	***-e**	-e (TV)	-e (TV)	-e (TV)	-e (TV)
Verbal							
Causative	***-s**	-s, -is	***-s**	-s, -is	-s, -is, -ʃ	-s, -is, -ʃ	-s, -is, -ʃ
Passive	***-am**	-am	***-am**	-am	-am	-am	-em
Reflexive	***-ʔ**	-ʔ, -d', -ad', -eed'	***-ʔ**	-ʔ, -i-t't', -ak'k'	-ʔ, -ak'k'	-ʔ, -id'	-ʔ, -ed'
Denominative	***ʔ(-C-C), *y(-C-C)**	-ʔ, -ad', -eed'	***ʔ(-C-C), *y(-C-C), *ʔ-k, *ʔ-t**	ʔ, -at, -it, -it't'	ʔ, -eek/-eeh, -it't'	ʔ, -id'	ʔ, -at, -ed'

Table 7: *Reconstructed HEC Morphemes and Their Reflexes*

The comparative distributions in **Table 7** show continuity in a core derivational inventory alongside language-specific changes in particular markers and combinations. Number morphology and nominal derivation exhibit both stable templates and variable exponents. The singulative suffix is reconstructable, but its surface form is conditioned: An earlier ***-ʃ** is generalized as ***-c** in many environments, with additional allomorphs determined by context. Reduplication of the final consonant recurs as a number-marking strategy in a subset of languages (Gedeo, Sidaama, Kambaata), and the plurative ***-uw-a** is best supported where it co-occurs with this reduplicative pattern (in Sidaama, Hadiyya, and Gedeo). Collective marking (***-ey-e**) is concentrated on adjectives (and a smaller set of nouns), with Burji showing distinct allomorphy. A further plurative ***-n** may underlie Burji **-na** and Sidaama **-aano**, but the comparative evidence remains limited and the reconstruction should be treated as provisional and mainly supported by fossilized evidence of ***-n** in the plural pronominals.

Inflectional morphology provides complementary evidence for both reconstruction and subgrouping. At the PHEC level, the gender consonants ***-k** (masculine) and ***-t** (feminine) are securely reconstructable given their uniform attestation across demonstratives, definiteness markers, and case morphology. Case marking is additionally inferred from terminal vowels, most transparently in pronominal and demonstrative paradigms, though substantial leveling in individual languages obscures earlier contrasts. The genitive exponent ***-ʔ** is more tentative: Direct segmental evidence is restricted, but it can be traced through conservative glottal retention and indirect reflexes (glottalization, metathesis, and accentual effects).

Verbal derivation shows a similarly template-driven organization. The denominative base is built from a verbalizer followed by an extension, and the template itself is reconstructable to PHEC even where individual complex forms are unevenly distributed. The passive is uniformly attested as ***-am** with a raised vowel in Gedeo **-em** and it is proposed to have received its vowel exponent via grammaticalization in the denominative construction. The causative suffix ***-s** is widespread, with a palatalized allomorph **-ʃ** in Kambaata, Sidaama, and Gedeo. Reflexive morphology offers one of the clearest cases for where disciplined reconstruction matters for subgrouping. A core reflexive ***-ʔ** is reconstructible to PHEC on the basis of its full HEC distribution and regular NHEC morpheme boundary alternations (including metathesis after sonorants and fusion after obstruents). In contrast, the **-(V)d'** variants found in Burji, Gedeo, and Sidaama are best treated as secondary, contact-driven changes (ultimately attributable to Oromo), and therefore

should not be projected to PHEC/PNHEC. Hadiyya and Kambaata innovated distinct reflexive exponents (**-i-t't'** and **-ak'k'**). This split between inherited suffixes, innovated structure, and a borrowed exponent underscores why diagnostic claims must rest on distribution, conditioning, and system fit rather than on isolated similarities.

The denominative analysis (§§4.2.4–4.2.5) establishes five core principles:

> Grammaticalization pathway: Support verbs [425] *ʔ 'to do' and [388] *y 'to say, do, be' grammaticalized into bound verbalizers, forming the denominative base (§4.2.4, 4.2.5.1, 4.2.5.2).
>
> Two-stage derivational template: Stage 1 combines nominal + verbalizer → verbalized stem; Stage 2 attaches extensions (agentive ***-t/*-k**, reflexive ***-ʔ**, passive ***-m**, causative ***-s**) → complex verbalized stem.
>
> Diachronic generalization: Complex denominatives (**ʔ+m** → ***-am, *y+s** → ***-is**) underwent functional expansion, being reanalyzed as general valency morphology (§4.2.5.9).
>
> Unified verbalizer system: Support verbs ***ʔ/*y** function identically across denominal (noun → verb) and valency (verb → verb) derivation.
>
> Differential grammaticalization: Language-specific variation (preservation in Burji, reanalysis in NHEC, contact-induced replacement) enables reconstruction of diachronic stages and supports subgrouping (Chapter 5).

A recurring conclusion across the verbal subsections is that diachronic change targets individual exponents more readily than it disrupts derivational templates. In NHEC, agentive extensions (***-t**, ***-k**) participate in stable template structures even as productivity shifts between competing markers and as some reflexes erode. The same template persistence is visible in the grammaticalization pathway from support verbs to verbalizers and onward to valency morphology, where complex denominative structures can fuse and be reanalyzed as synchronic passive/causative/reflexive morphology.

Table 8 presents a systematic reconstruction and a condensed comparative overview of the PHEC and PNHEC personal pronoun paradigm across three cases (nominative, accusative, possessive) and all person/number

distinctions. The paradigm illustrates several key morphological developments: (i) metathesis of possessive glottal stems, which serves as a diagnostic PNHEC innovation; (ii) paradigm leveling in the 2pl series, where glottal insertion (***ati-n-** → ***aʔ-n-**) and restructuring create analogical regularization; and (iii) stability of 3rd plural forms, which maintain PHEC shapes through both proto-stages. Detailed analysis of these reconstructions is provided in §4.1.6 (personal pronouns) and §4.1.4.2 (possessive/genitive pronouns).

Table 8: *Reconstructed HEC Pronouns: Nominative, Accusative, and Possessive*

1sg	**[2]**	**nom**	**[103]**	**acc**	**[135]**	**poss**
PHEC	***an-i**	‘I’	***e-y**	‘me’	***ʔ-y**	‘my’
PNHEC	***ani**	‘I’	***ey**	‘me’	***ʔy**	‘my’
1pl	**[63]**		**[197]**		**[188]**	
PHEC	***naʔ-**	‘we’	***n-y**	‘us’	***ni-ʔ-**	‘our’
PNHEC	***naʔ-**	‘we’	***n-y**	‘us’	***niʔ-**	‘our’
2pl	**[104]**		**[196]**		**[187]**	
PHEC	***ati-n-**	‘you’ (pl)	***ki-n-**	‘you’ (pl)	***kin-ʔ**	‘your’ (pl)
PNHEC	***aʔ-n-**	‘you’ (pl)	***kiʔn-**	‘you’ (pl)	***kin-ʔ-** → ***kiʔn-**	‘your’ (pl)
2sg	**[27]**		**[257]**		**[253]**	
PHEC	***at-i**	‘you’ (sg)	***ke**	‘you’ (sg)	***ki-ʔ-**	‘your’ (sg)
PNHEC	***at-i**	‘you’ (sg)	***ke**	‘you’ (sg)	***ki-ʔ, *-ʔ-ki**	‘your’ (sg)
3sg.f	**[6]**		**[151]**		**[366]**	
PHEC	***is-i, *is-y**	‘she’	***is-e, *is-é**	‘her’	***ise-ʔ-**	‘her’
PNHEC	***is-i**	‘she’	***is-e**	‘her’	***ise-ʔ-**	‘her’
3sg.m	**[7]**		**[292]**		**[279]**	
PHEC	***is-u, *is-w**	‘he’	***is-o, *is-ó**	‘him’	***isa-ʔ-**	‘his’
PNHEC	***is-u**	‘he’	***is-ó**	‘him’	***isi-ʔ-**	‘his’
3pl	**[11]**		**[86]**		**[318]**	
PHEC	***is-n-**	‘they’	***is-n-**	‘them’	***isn-ʔ-**	‘their’
PNHEC	***is-n-**	‘they’	***is-n-**	‘them’	***isn-ʔ-**	‘their’

The pronominal paradigm summarized in **Table *8*** supports a compact set of PHEC reconstructions that integrate case, number, and possession. Third-person forms share a stem **is-* across nominative and accusative, with case-marking for feminine (***-i** nominative vs. ***-e** accusative) and for masculine (***-u** nominative vs. ***-o** accusative). In the plural, a segmentable plurative ***-n** is recoverable in several forms, while possessive morphology allows a genitive ***-ʔ** to be inferred even where direct segmental reflexes are reduced.

The morphological reconstruction demonstrates that HEC synchronic diversity arises from three interacting processes: (a) regular sound change applying consistently across morpheme boundaries (Ch. 3); (b) analogical paradigm

leveling creating language-specific irregularities; and (c) grammaticalization pathways differentially advanced across daughter languages. The reconstructed morphemes summarized in **Table 7** and **Table 8** provide diagnostic morphology for the HEC subgrouping developed in Chapter 5. Burji reflects archaic retention in several domains, while NHEC exhibits shared innovations such as including metathesis in the pronominal system, causative palatalization, and changes in agentive extensions, which define the northern branch. Across nominal and verbal morphology, the same grammaticalization trajectory recurs (support verb → verbalizer + extension → valency marker), consistent with a template-based derivational structure inherited from PHEC. Future work should target the weakest points in the current reconstruction – i.e., especially genitive alignment and case morphology – to test whether they align with the same branch-defining innovations or reflect independent innovations and contact.

5 Conclusions

This chapter synthesizes the results of Chapters 1–5 and states what they jointly imply for reconstructing Highland East Cushitic. It also separates inherited structures from later changes, where the comparative evidence requires caution. Phonological reconstruction and morphophonological reconstruction are treated here as one analytical system: Segment inventories, prosody, phonotactics, and boundary alternations are interpreted together because the strongest internal diagnostics for reconstruction often come from morphologically conditioned environments.

5.1 What the Study Has Established

5.1.1 Phonological Reconstruction

Chapter 2 reconstructs the phonological system of PHEC on the basis of regular sound correspondences across the five HEC languages (Burji, Hadiyya, Kambaata, Sidaama, and Gedeo). The reconstruction is explicit about degrees of confidence as proto-phonemes are evaluated according to the regularity of the correspondences, the breadth of the supporting examples, and the strength of the comparative evidence.

The reconstruction yields a PHEC inventory with a five-vowel system and contrastive quantity alongside relatively symmetrical consonant subsystems. The consonantal system includes both pulmonic and non-pulmonic series, and the analysis highlights that not all complex correspondence patterns are equally diagnostic. The stops, nasals, and most vowels tend to show the most reliable regularities, whereas sibilant behavior is more complex, interacting with regular sound change, harmony patterns, and (in some cases) areal influence.

Chapter 2 establishes the basic phonotactic and prosodic frame. PHEC accent is reconstructed as weight-sensitive, with later changes in terminal vowels contributing to systematic differences in accent placement across the daughter languages. The result is a coherent picture in which quantity, syllable structure, and prominence are not independent components, but mutually constraining dimensions of the same reconstructed system.

Chapter 3 continues the reconstruction by shifting the evidential focus from lexemes and segmental correspondences to alternations at morphological boundaries. It synthesizes the major morphophonological processes, summarizes their distribution across the HEC languages, and evaluates which

processes are reconstructable to PHEC and which are better treated as later innovations (including changes that may reflect contact).

Two structural domains are especially prominent. First, boundary-driven repairs in stem+suffix combinations frequently create an additional syllable to resolve impermissible consonant sequences, most commonly through vowel or glide epenthesis (including interconsonantal vowel/intervocalic glide epenthesis that must be distinguished from support-verb morphology where relevant). Second, hiatus repair is resolved through glide insertion (***-y-**, with ***-w-** as a conditioned realization near back vowels) or, in a more restricted subset of languages, by glottal stop insertion.

The comparative distribution of these repairs matters. Interconsonantal epenthesis is broadly productive, but the same cognate set does not generally show uniform epenthesis across languages, which limits the reconstructibility of many surface insertions. On the evidence assembled in Chapter 3, only ***-y-** (together with its conditioned realizations) is proposed as a secure PHEC epenthetic segment, while other consonantal insertions are treated as later, language-specific (or contact-induced) changes unless future comparison establishes regular correspondences.

5.1.2 Morphological Reconstruction

Chapter 4 4extends the reconstruction to nominal and verbal morphology. On the basis of reconstructed morphemes and their reflexes across the HEC languages, it argues that the historical change of HEC morphology reflects a complex interaction between inherited morphemes and later changes that may include contact-induced change.

The chapter's primary contribution to the conclusion synthesis is twofold: It consolidates a cross-linguistic inventory of reconstructable morphemes (with explicit gaps and uneven attestations), and it provides a comparative basis for distinguishing what can be projected to PHEC/PNHEC from what is better analyzed as later restructuring. Where the morphological evidence bears on subgrouping or on contact histories, those implications are taken up separately in §§5.3–5.4.

5.2 Concluding Phonology Synthesis

This section synthesizes the phonological and alternations at morpheme boundaries results of Chapters 1–3 as a single system (segmental inventories, prosody, phonotactics, and boundary processes). Chapter 4 supplies the

morphological structures (affixes and paradigmatic environments) that make the boundary alternations in Chapter 3 interpretable as structured evidence rather than isolated irregularities to support the lexical reconstruction. Subgrouping arguments and a discussion on language contact are treated in §5.3.

5.2.1 Phonological System

Chapters 1–3 motivate a simple methodological point with wide consequences for reconstruction: In HEC, the phonological system is most clearly exposed where morphology forces difficult sequences. Segment inventories and prosodic generalizations are not only visible in regular correspondences but also in the ways stems and affixes interact.

The key advantage is interpretive rather than merely descriptive. Boundary alternations show which constraints are active (and stable enough to be reconstructed), and they provide a principled way to diagnose non-initial consonants and consonant clusters that are otherwise obscured by reduction, fusion, metathesis, and variable surface repairs. The morphological analysis in Chapter 4 is crucial here because once the relevant affixes and paradigmatic contexts are identified, seemingly disparate alternations resolve into a small set of recurring strategies.

5.2.2 Segmental Profile of PHEC

At the segmental level, the vowel system is both compact and stable: A five-vowel inventory with a phonemic quantity contrast (treated moraically) is recoverable across the family. The conclusion that matters for later analysis is that vowel quantity and terminal-vowel behavior jointly structure prosody and many surface word forms.

Terminal vowels are a shared property across HEC, but their realization varies by language and morphosyntactic context (case/gender marking) and by diachronic reduction (apocope and devoicing in some cases). This diversity in realizations helps explain why accent targets shift, why word edges are frequent sites of erosion, and why boundary contexts become especially informative.

The reconstructed 21-consonant system is comparably stable in its overall size and typological profile, including both pulmonic obstruents/sonorants and a non-pulmonic series. The analytical lesson from Chapters 1–3 is that consonantal reconstruction is often easiest in positions where the surface is

least stable. Unlike in initial position, the non-initial sounds and clusters are precisely the target where boundary alternations, distributional restrictions, and repair choices provide decisive diagnostics.

5.2.3 Phonotactics and Prosody

Phonotactically, the system can be summarized as a restrictive root template: (C)V(V)(C)(C), with a strong preference for open syllables and a ban on initial clusters. Where clusters occur, they are limited and shaped by sonority sequencing. The broader generalization is that “complexity” is managed rather than eliminated. The HEC family permits certain coda and medial configurations, but it regulates them tightly, which in turn makes repairs predictable at morpheme boundaries.

The prosodic system is best understood as weight-sensitive. Prominence aligns with heavy syllables (those with double vowels or closed ones), but surface placement is frequently restricted to the ultimate/penultimate syllable. The family-wide pattern supports a diachronic scenario in which an earlier “ultimate heavy” target is variably reanalyzed as penultimate prominence when terminal vowels reduce or lose the final consonant needed to bear accent.

5.2.4 Boundary Alternations

Boundary alternations are the meeting point between the abstract phonological system (what sequences are permitted) and the morphological system (what sequences are demanded). Across the languages, the most revealing patterns concentrate in stem + suffix concatenation and in vowel-vowel adjacency (hiatus), precisely where the syllable template and cluster constraints would otherwise be violated.

> Assimilation provides the most direct evidence for active feature pressures at boundaries, and it frequently targets obstruents.
>
> Fusion and gemination show how boundary configurations can be adapted to fit into segmental outcomes (including glottalization in some processes), turning morphological structure into reconstructable phonological evidence.
>
> Palatalization emphasizes the close relationship between segmental modification and the vocalic environment influenced by affixation.

Insertion (epenthesis) and deletion (elision) represent complementary responses to constraint pressure – either add segments to make a form pronounceable or remove material that creates an illicit configuration. However, not every inserted segment is reconstructable; the comparative signal most strongly supports an epenthetic ***-y-*** at the PHEC level, while many other insertion choices appear to be later, language-specific solutions.

Metathesis is common in NHEC and rare in Burji. It is revealing where it clusters in the same prosodic and phonotactic "problem zones", again tying surface diversity to a small set of underlying constraints.

Chapters 1–3 support a conservative principle that treats boundary alternations first as evidence about constraints and the permissible repair processes, and only second as evidence for particular segments.

5.2.5 HEC Sound Classes

The boundary evidence motivates a high-level sound classification that is useful for both description and explanation. The obstruents behave as the main target of feature change, while sonorants behave as the main resource for making forms well-formed.

Obstruents are preferentially affected by feature-changing processes (especially assimilation) and participate readily in boundary outcomes like fusion and gemination. Sonorants, by contrast, appear repeatedly as "phonotactic mediators", so they are favored in epenthesis to satisfy sequencing preferences and to resolve illicit clusters, and they participate in metathesis patterns that improve syllable structure in morphological formations.

Table 9: *Major Sound Classes and Boundary Processes*

Major Class	Minor Class	Assim.	Fusion & gemin.	Metath.	Inser.
Obstruent	Stop	✓	✓		(-ʔ-)
	Non-pulmonic	✓	✓		
	Fricative	✓	✓		
Sonorant	Lateral & Trill	✓		✓	✓
	Nasal	✓		✓	✓
	Glide				✓

Table 10 organizes the consonant inventory by place and manner of articulation within these two major classes. It serves as a compact reference point for the segment sets implicated in assimilation, lenition, and cluster-repair patterns in **Table 9**.

Table 10: *PHEC Consonant Inventory Organized by Class (place × manner)*

		Labial	Alveolar	Post-alveo	Velar	Phary.	Glottal
Obst.	Stop	**b**	**t, d**		**k, g**		**ʔ**
	Non-pulmonic		**t'**	**c'**	**k'**		
	Fricative	**f**	**s, z**	**ʃ**		**ħ**	**h**
Son.	Lateral & Trill		**l, r**				
	Nasal	**m**	**n**				
	Glide	**w**		**y**			

The combined evidence of Chapters 1–3 supports a coherent diachronic picture in which a constrained phonotactic system is maintained through a set of recurrent boundary strategies. The value of the morphophonological record is that it makes those constraints visible and comparable across languages, even where non-initial sounds are unstable on the surface. This synthesis provides the phonological and morphological background needed for the next parts of Chapter 5, where shared innovations are evaluated for subgrouping and where contact-induced patterns are separated from broadly inherited tendencies.

5.3 Subgrouping Evidence

Highland East Cushitic (HEC) is well established as a genetic unit within the East Cushitic family in the comparative literature. As Sasse (1982, p. 9) recalls, "Highland East Cushitic (HEC), a term coined by Bender (1971) to comprise the languages Sidamo, Darasa (Gedeo), Kambata (with two parallel dialects: Timbaro and Alaba), Hadiyya and Burji… grew out of the traditional 'Eastern Sidama group'." In the Language Survey of Ethiopia, "Fleming follows Moreno (1940) in treating HEC as a subgroup in East Cushitic" (Hudson, 1976, p. 246). Hudson (2006, pp. 295–296) further synthesizes the typological profile that characterizes the HEC languages. Here, I advance an innovation-based classification, based on the enhanced coverage afforded by the updated comparative database (§1.3) and detailed reconstruction in the previous chapters.

The subgrouping argument in this study is based on converging evidence across domains rather than on any single diagnostic parameter. The most heavily weighted criteria are shared innovations with restricted distributions, especially when they involve bound morphology or productive alternations (because such patterns are less likely to be independently replicated across languages than isolated lexical similarities).

Accordingly, this chapter distinguishes three kinds of comparative signal:

> **Shared inherited background** (reconstructable to PHEC): processes and distributions that are present across Burji and the NHEC branch and therefore define the diagnostics against which later changes must be evaluated. Measured exceptions supported by Cushitic evidence allow reconstruction to PHEC despite restricted distribution.
>
> **Branch-defining innovations** (at the PNHEC level or below): processes whose distribution aligns with a coherent subgroup and which are absent (or different) in Burji.
>
> **Language-specific changes**: innovations that concentrate in a single language and therefore clarify divergence without themselves defining an internal subgroup.

Because many HEC patterns are morphologically conditioned, the subgrouping logic is explicitly cumulative: phonological innovations (Chapter 2) are interpreted together with morpheme boundary alternations (Chapter 3) and shared morphological markers (Chapter 4), while contact-driven irregularities are treated separately (§5.4).

The lexical evidence is integrated on the same comparative terms. The East Cushitic Database (§1.3) links reconstructions and correspondence judgments to explicit cognate sets and source documentation via stable IDs, so that each claim can be checked against the data upon which it was made. The reconstruction practice also enforces a minimum-attestation threshold: a proto-form is reconstructed only when a cognate set is supported by at least three daughter language cognates within the relevant scope. This has a direct subgrouping consequence. A set attested in Burji plus two NHEC languages can license a PHEC reconstruction without licensing a PNHEC reconstruction; absence at the PNHEC level therefore reflects insufficient internal evidence within the available dataset, not historical absence. A substantial portion of lexical reconstructions are identical at the PHEC and PNHEC levels,

while PNHEC also shows a set of reconstructable lexemes and changes that are not found in Burji.

5.3.1 Phonological Diagnostics

The phonological reconstruction in Chapter 2 establishes that Burji shows the earliest and most regular divergences from the rest of the HEC languages, while the remaining languages share a set of changes that Burji lacks. This asymmetry is the first subgrouping result, supporting an initial split between Burji and a Northern Highland East Cushitic (NHEC) branch.

Chapter 2 also names concrete innovations that help diagnose this split. Three examples that are stated explicitly in the chapter's synthesis are:

> **Shared NHEC innovation:** loss of initial ***k'** and ***ħ**, treated as part of the shared innovation set supporting NHEC unity.
>
> **Burji innovation:** a Burji-specific stop-sibilant metathesis pattern (inherited $*\mathbf{C}_{[\text{velar}]}\mathbf{s}$ sequence metathesizes in Burji), identified as a distinct innovation of the Burji line after its split from what would become PNHEC.
>
> **Systematic Burji vs. NHEC divergence:** in certain cognate sets, Burji consistently shows **d'** where NHEC languages show **t'**, a recurrent correspondence asymmetry that aligns with the early separation of Burji.

At this level of the synthesis, the critical point is methodological. A subgrouping claim is strongest when multiple independent innovations align with the same split. For this reason, the phonological evidence from Chapter 2 is not treated in isolation here, but taken together with the alternations at morpheme boundaries and morphological innovations.

Chapter 3 provides especially direct subgrouping evidence because morphophonological processes are (a) paradigmatically supported, (b) structurally conditioned, and (c) often distributionally restricted in ways that map neatly onto internal branches.

Table 11: *Morphophonological Subgrouping Diagnostics*

Process/Feature	**Distribution**	**Proto-Level interpretation**	**Subgrouping implication**
Full **-t** assimilation to preceding **C**	Had, Kam, Sid, Ged (NOT Burji)	PNHEC innovation	Supports Burji/NHEC split
t → **c/ʃ**, **k** → **ʃ** palatalization before **i**	Burji only	Burji innovation	Burji-specific
Obstruent + **ʔ** → glottalized geminate (fusion)	Had, Kam, Sid, Ged (NOT Burji)	PNHEC innovation	Supports Burji/NHEC split
Stop + **s** → metathesis	Burji only	Burji innovation	Burji-specific
Sonorant + ʔ → metathesis	Had, Kam, Sid, Ged (NOT Burji)	PNHEC innovation	Supports Burji/NHEC split

Several boundary-driven patterns are shared across Burji and the NHEC branch and can therefore be reconstructed at the PHEC level with relatively high confidence. These include: assimilation of the plurative pronominal suffix to a preceding stem-final sonorant, affrication affecting a nominal derivational suffix, intervocalic lenition of stops, and common repair strategies such as intervocalic/interconsonantal epenthesis (with an epenthetic ***-y-*** established as secure). These shared properties define inherited diagnostics: Similarities in these domains do not, by themselves, constitute evidence for a PNHEC subgroup.

Against that inherited baseline, several morphophonological processes show an exclusive NHEC distribution and thus function as PNHEC diagnostics (or narrower subbranch innovations) that contribute directly to internal classification. In particular, full assimilation of the pronominal suffix ***-t*** after obstruents, the fusion of an obstruent with a following glottal stop yielding glottalized geminates, and metathesis in sonorant + glottal stop sequences are restricted to NHEC and absent in Burji. These innovations jointly support the Burji vs. NHEC split and reinforce the divergence patterns established on phonological grounds.

Burji, conversely, shows morpheme boundary alternations not shared with NHEC. The most notable are the palatalized reflexes before ***i*** (or ***y***) environments and the unique stop + sibilant metathesis pattern mentioned above. These patterns confirm that Burji diverged before the NHEC innovations took place.

5.3.2 Morphological Diagnostics

Morphological innovations are among the strongest types of evidence for subgrouping because shared markers and paradigmatic structures are less likely to be replicated independently across languages. Chapter 4 demonstrates that morphology is also a locus for contact-induced change. In this synthesis, Chapter 5 therefore distinguishes (a) inherited morphology that defines the PHEC baseline from (b) branch-specific innovations and (c) contact-driven lookalikes.

5.3.2.1 Nominal Morphology

In nominal morphology, Chapter 4 treats gender consonants as among the most secure reconstructions. PHEC ***-k** (masculine) and ***-t** (feminine) are reconstructable across demonstratives and definiteness markers, and are attested in gender and case morphology. Their comparative stability and inherited background provide a stable baseline against which later developments can be evaluated rather than serving as a diagnostic innovation by themselves.

The singulative and deverbative ***-ʃ** is also widely attested throughout HEC with diverse allomorphy. Therefore, it is a secure PHEC reconstruction. The only distinction between the singulative and the deverbative is in their templates: nominal stem + ***-ʃ** → singulative, while verbalized stem + ***-ʃ** → deverbative.

The remaining exponents distribute within smaller groups comprising 2–3 languages: Sidaama, Gedeo, Burji (collective ***-ey-e**), Hadiyya, Sidaama, Gedeo (plurative ***-uw-a**). However, their reconstruction to PHEC still qualifies as reasonable based on wider Cushitic evidence. The same applies to the stem-final reduplication plurative (***C$_i$~C$_i$-a**) found in all HEC and other Cushitic languages, but not in Burji.

The genitive exponent ***-ʔ** is reconstructed to PHEC with arguable certainty, with clearer segmental evidence in some languages and indirect reflexes in others (e.g., glottalization or accentual effects after glottal loss).

5.3.2.2 Pronominal Morphology

The pronominal paradigm provisionally reconstructed in **Table 8** (§4.3) reveals three defining structural properties. First, suppletion is confined to first and second persons (distinct roots (**an-*, **at-*, **naʔ-*, **ati-*)), contrasting with

the third person built on ***is-**. Second, case marking strategies differ by subsystem: The nominative employs vowel suffixes (***-i, *-u**), the accusative uses suppletive forms (**ke, *is-o, *is-e*) or vowels, while the possessive relies on the genitive glottal (**ʔ-y*, **ni-ʔ-*, **kin-ʔ-*). Third, metathesis affects possessive glottal stems in the PHEC → PNHEC transition (e.g., PHEC **kin-ʔ-* becomes PNHEC **kiʔn-* in 2pl), serving as a diagnostic PNHEC innovation. The 3pl possessive **isn-ʔ-* resists metathesis due to consonant clustering, triggering vowel insertion across the reflexes.

These morphological patterns reveal that the pronominal paradigm participated fully in systematic sound changes, differentiating PHEC from PNHEC (Ch. 3). Metathesis of possessive glottals represents paradigm-specific conditioning (glottal + nasal sequences), while 3pl cluster resolution demonstrates phonotactic repair. The interplay between phonological regularity (the metathesis rule applies) and morphological analogy (paradigm leveling in 2pl PHEC **ati-n-*, PNHEC **aʔ-n-*) provides evidence for the integrated morphophonological reconstruction undertaken in this study.

5.3.2.3 Reflexive Morphology

Reflexive morphology is diagnostic for subgrouping because Section 4.2.3 explicitly separates an inherited exponent from later, distributionally restricted developments. A core reflexive marker ***-ʔ** is securely reconstructable to PHEC on the basis of broad distribution and regular morpheme boundary alternations.

Within NHEC, §4.2.3 treats the reflexive alternations at morpeheme boundaries as part of a regular conditioning profile, i.e., metathesis after sonorant-final stems and fusion after obstruent-final stems (yielding glottalized geminate outcomes). In addition, complex reflexive formations such as **-ak'k'** / **-it't'** are described as restricted to Kambaata and Hadiyya and are therefore candidates for narrower internal NHEC innovations (cf. §4.2.5.9).

With regard to the contact discussion in §5.4, see also §4.2.3.4, which reports reflexive variants involving **-(V)d'** in Burji, Gedeo, and Sidaama, and treats them as borrowed from Oromo and not reconstructable to PHEC or PNHEC. These forms should be excluded from subgrouping arguments and treated instead as a contact signal.

5.3.2.4 Verbal Derivation

The verbal subsections (§§4.2.4–4.2.5) emphasize a recurring diachronic asymmetry: Change often targets individual exponents while preserving derivational templates (slot structure). This supports reconstructing the verbal template architecture to PHEC, even where particular extensions are unevenly distributed. Section 4.2.5 also invokes a grammaticalization trajectory (support verb → verbalizer → valency marker), which provides a coherent explanation for why template structure can remain stable while particular formations are replaced or reanalyzed.

For subgrouping, the template architecture represents the PHEC baseline rather than diagnostic innovations. Diagnostic potential lies in restricted lower-level innovations, in particular exponents or allomorphs such as **-it't'** in Kambaata and Hadiyya.

Table 12: *Morphological Subgrouping Diagnostics*

Exponent/structure	**Distribution**	**Proto-level interpretation**
Passive ***-am**	All HEC	PHEC (secure)
Causative ***-s**	All HEC	PHEC (secure)
Causative **-ʃ**	Kam, Sid, Ged	Innovation (narrow)
Gender ***-k** (m), ***-t** (f)	All HEC	PHEC (secure)
Genitive ***-ʔ**	Had + HEC (indirect)	PHEC (arguable)
Plurative ***C_l~C_l-a**	NHEC	PHEC (reasonable)
Plurative ***-uw-a**	Had, Sid, Ged	PHEC (reasonable)
Collective ***-ey-e**	Sid, Ged, Burji	PHEC (reasonable)
Singulative / Deverbative ***-ʃ**	All HEC	PHEC (secure)
Reflexive *-ʔ	All HEC	PHEC (secure)
Reflexives **-ak'k' / -it't'**	Had, Kam	Innovation (narrow)
Reflexive **-Vd'**	Burji, Ged, Sid	Contact-induced
Verbal templates (exponent turnover)	All HEC	PHEC (reasonable)

5.3.3 Subgrouping and Revised EC Labels

Chapters 1–4 establish an initial split between Burji and an NHEC branch comprising Hadiyya, Kambaata, Sidaama, and Gedeo. The evidence is cumulative: Phonological divergence patterns are reinforced by multiple NHEC-specific morphophonological innovations and by Burji-specific developments that concentrate outside the NHEC distribution.

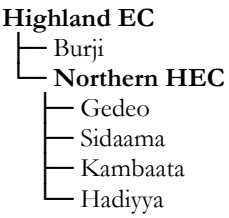


Figure 2: Highland East Cushitic Family Tree (introduced in §1.1.3)

More fine-grained internal subgrouping within NHEC requires careful filtering for contact effects (§5.4) and a fuller integration of the morphological distributions. For this reason, the present synthesis treats the Burji vs. NHEC split as the primary subgrouping result, leaving more detailed internal branching to a subsequent analytical step. This revises and refines Hudson's (2007, 1989, 1981) HEC group relationships, where Hadiyya-Kambaata and Sidaama-Gedeo share common nodes.

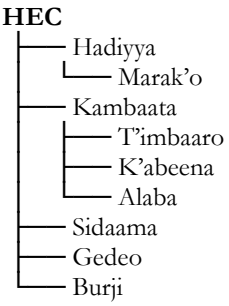


Figure 5: *Hudson's Family Vine Tree (2007, 1989, 1981)*

Hudson's (1981) "family vine" envisions the HEC languages as a gradient continuum or chain. In the family vine, Burji is the closest to Gedeo, Gedeo to Sidaama, Sidaama to Kambaata, and Kambaata to Hadiyya based on a quantitative count of shared grammatical features from a comparative survey (Hudson, 1976). This vine-like arrangement, illustrated via relative strengths

of shared traits (e.g., phonological, morphological, lexical), emphasizes dialect geography and wave-like diffusion over discrete branching, while acknowledging the challenges of distinguishing innovations from retentions or borrowings. Thus, Hudson's quantitative enumeration of features treats all features equally without prioritizing innovations. The current study refines this by focusing on diagnostics, confidence evaluation, and explicit contact diagnostics, while excluding probable areal diffusions such as **-(V)d'** in Burji, Sidaama, and Gedeo. By layering contact filtering, the current study aligns the vine's gradient insights with a genealogical hierarchy, where PNHEC emerges as a robust node.

Lexically, Hadiyya-Kambaata and Sidaama-Gedeo pairs share the most cognates (confirming Hudson's subgrouping), followed by Hadiyya-Kambaata-Sidaama. However, no solid shared innovations support internal subgrouping of the four NHEC languages: Hadiyya, Kambaata, Sidaama, and Gedeo. Contact signatures provided false signals for many of Hudson's shared features used in his subgrouping. NHEC-internal contact likely contributed to the shared lexicon in these groups, while the signals Burji and Gedeo share derive from their common donor, Oromo.

In sum, HEC forms a secure genealogical unit defined by a cross-domain bundle of innovations (Ch. 2–4). Internally, the primary split separates Burji from PNHEC, supported by a number of innovation bundles and, independently, by loan distributions. Within PNHEC, there is no signal strong enough to allow the positing of a reliable subgrouping based on shared innovations. Hadiyya and Kambaata preserve relatively conservative profiles that aid the reconstruction.

I use revised traditional subgroup labels in this study for four EC subgroups. Where traditional labels use a single language name as a cover term for a subgroup node, I adopt descriptive, geographically anchored terms for clarity (**Northern EC** for Saho-Afar; **Central EC** for Oromoid; **Sagan** for Konsoid, **Southern EC** for Yaaku-Dullay). I also propose these terms as a practical alternative for future comparative work, since they make the intended scope of each node more transparent while avoiding unintended asymmetries by remaining strictly terminological. Other than the new "Northern HEC" designation, these labels function as descriptive terms and do not imply new subgrouping claims.

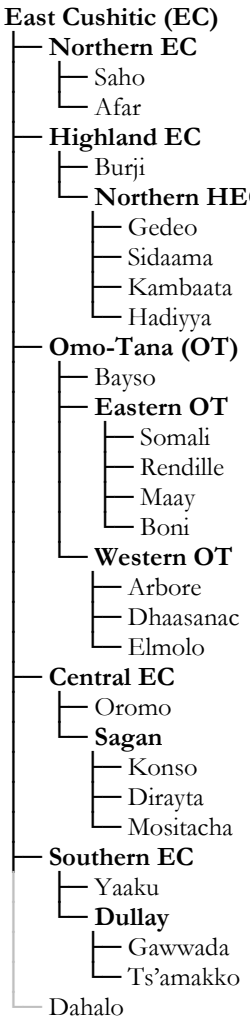


Figure 6: *Revised East Cushitic Subgroups Labels*

This study does not posit any arguments against Lowland East Cushitic, and even proposes an arguable diagnosis for the Lowland East Cushitic node based on the non-pulmonic evidence discussed in §2.1.2 pending further research.

The next section (5.4) turns to the contact ecology and the layered borrowing history (involving Oromo, Amharic, and Omotic) that shaped the lexicon and left recognizable donor signatures without overturning the genealogical signal.

5.4 Language Contact and Diffusion

Language contact is treated in this study as a pervasive historical factor rather than as an occasional anomaly. Interaction with neighboring languages, especially Oromo and Omotic, has left detectable traces across HEC, though the degree and region of influence vary by language and by domain.

On general comparative grounds, the most expected contact areas are (a) the lexicon (loanwords and loan translations), (b) distributional clustering of variants by geography rather than by inheritance, and (c) irregular reflexes or double outcomes that resist a regular correspondence explanation.

The most direct contact evidence is in the lexicon, where borrowed forms enter without exhibiting inherited correspondences and introduce segments or clusters that are otherwise marginal. For this reason, lexical distributions are interpreted cautiously when used as evidence of subgrouping. The discussion in §5.3, therefore, treats lexical signals as strongest when they align with innovation-based evidence from phonology and morphology.

Debates about Burji's position also underscore why contact must be modeled explicitly in subgrouping. Sasse (1982, pp. 9–10) frames the relationship of Burji to HEC as a long-standing classificatory problem and notes that Burji shows, in some respects, stronger affinities to Lowland East Cushitic than other languages classified as HEC. Hudson (1981, p. 116) likewise treats language contact as a potential contributor to Burji's divergence, while emphasizing that Burji's grammatical divergence is not straightforwardly attributable to borrowing. In this study, such observations motivate contact filtering and conservative interpretation of lexical distributions rather than any relaxation of innovation-based subgrouping criteria.

Regarding contact, the literature has repeatedly noted substantial Oromo influence on parts of the HEC lexicon, especially in Burji and Gedeo. Hudson (1989, p. 8) describes all five HEC languages as showing apparent borrowings from Oromo and Amharic, with Burji and Gedeo standing out for their especially strong Oromo lexical influence. Hudson (1976, p. 276) similarly observes: "In Burji and Darasa (more so in Darasa) [Gedeo], the lexicon is considerably penetrated by Oromo loans. But for neither language should these loanwords inspire doubts about its genetic affiliation". This study also confirms that Burji's and Gedeo's divergence is due to heavy contact with a (Lowland) East Cushitic language, and the evidence comes from Oromo. Contact does not necessarily appear as a broad disruption of correspondence

structure. It more often surfaces as restricted irregularity, geographically clustered variants, or outcomes that sit outside the inherited system and therefore require a donor hypothesis. This matters for interpretation because the same surface irregularity can reflect borrowing, diffusion, or the coexistence of inherited reflexes beside borrowed ones.

Morphology is integrated into this discussion for two reasons. First, borrowed segments (e.g., **d'**) can be amplified through morphological analogy, increasing their apparent regularity. Second, where morphological restructuring aligns with known contact zones, it must be evaluated cautiously before being treated as evidence for subgrouping. Chapter 4 provides explicit examples of this logic in the reflexive domain, where **-(V)d'** variants are treated as Oromo-induced developments and are excluded from proto-level reconstruction.

The working diagnostic for separating contact from inheritance follows the evidential logic stated in §1.5. In practice, the analysis privileges: (a) regular correspondences within HEC to establish inherited structure; (b) distributional irregularities as potential indicators of contact; (c) wider East Cushitic comparison to establish what counts as regular before positing a contact source; and (d) areal comparison where it clarifies direction and plausibility of borrowing. For example, comparing the non-pulmonic PHEC ***t'** and PHEC ***c'** from Chapter 2 (§2.1.2), their similarities include (a) being proto-phonemes reconstructed as inherited, (b) showing regular correspondences across HEC, (c) showing position-independent retention (initials/non-initials), (d) having complications from contact (Oromo, Amharic), and (e) having correspondences in the rest of EC. The key differences, on the other hand, are summarized as follows:

Feature	PHEC *t’	PHEC *c’
Primary status	Inherited non-pulmonic	Inherited post-alveolar non-pulmonic
Contact complications	Irregular **d’** and **c’** reflexes in subset	Amharic loanwords with regular **c’**
Oromo influence	Causes irregular **d’** and **c’** in Burji, Gedeo, Sidaama, and “double reflexes”	Causes “double reflexes”
EC correspondences	Strong, widespread	More parallels, mostly regular but less systematic data
Non-Cushitic parallels	More parallels	Fewer parallels
Distribution	Appears more common	Appears less common (fewer examples cited)

The crucial difference is the irregular vs. regular **c’** reflexes. The irregular **c’** (from ***t’**) is contact-induced, limited to Burji, Gedeo, and Sidaama, and found on specific lexemes. In contrast, the regular **c’** (from ***c’**) is retained, found in all five HEC languages, and has regular correspondences. This distinction justifies reconstructing both ***t’** and ***c’** as separate proto-phonemes.

Where a pattern is fully regular within HEC and supported by substantial comparative evidence, the analysis generally does not attempt to partition the outcome into “inheritance” versus “contact” components, because any ancient influence may have been fully regularized. In such cases, parsimony favors analysis as inherited unless external evidence provides a stronger reason to posit contact. Where irregularity is present, however, contact remains the preferred hypothesis and is incorporated into confidence evaluation rather than excluded by default.

5.5 Methodological Implications

Across the reconstruction (Ch. 2–4), claims are not treated as equally secure by default. The analysis follows the probability-based evaluation framework established in §1.5 and assigns confidence evaluations to reconstructions based on (a) the regularity of the correspondence patterns, (b) the number and distribution of supporting examples, and (c) the strength of the comparative evidence.

In practice, this means that the study prioritizes convergence of evidence from multiple sources: regular correspondence sets, distributional constraints, and morpheme boundary alternations that expose underlying structure. When correspondence patterns are complicated by harmony, positional

conditioning, or areal convergence, the evaluation framework makes that uncertainty explicit rather than forcing a categorical reconstruction.

Although contact and inheritance are treated as a continuum in the history of HEC (rather than binary alternatives), reconstructions are assessed with an attention to distinguishing inherited structure from later innovation. This chapter maintains this separation: §5.2 synthesizes the phonological and morphophonological system, while §§5.3–5.4 develop the subgrouping and contact interpretations.

Boundary alternations provide a different kind of comparative evidence than lexical correspondences do. They are valuable precisely because they are systematic, productive, and tightly conditioned by morphological concatenation. In the HEC data, stem + suffix environments repeatedly trigger repairs that would be hard to motivate from citation forms alone, especially when the surface outcomes differ across languages.

Two points are vital for reconstruction. First, boundary-driven syllable creation (often via vowel/glide material) and stem-final consonant alternations clarify what counts as an impermissible cluster in each system and show which repair strategies are productive. Second, hiatus and alternation-repair identify a limited set of recurrent solutions whose distribution is itself comparative evidence, e.g., glide insertion patterns (***-y-**, with ***-w-** as a conditioned realization near back vowels) contrast with more restricted glottal stop insertion. Boundary alternations also force the analysis to separate proto-level structure from later changes. Some starred input patterns used in summaries (§3.5) are best treated as schematic or intermediate stages rather than direct proto-forms, because subsequent assimilation and other boundary processes may obscure the underlying composition.

A recurring limitation in the reconstruction is that the absence of an alternation is rarely decisive evidence for absence at the proto-stage. Attestation gaps, uneven descriptions, and differences in analysis all distort the surface distribution of boundary processes. For this reason, negative evidence is treated cautiously throughout the study.

Reconstruction, therefore, proceeds by ranking hypotheses rather than maximizing proto-inventory. When a process is broadly productive but does not align across cognate sets, it is more plausibly a later innovation (including contact-induced developments) than a retained proto-level feature. This

conservative stance keeps the reconstruction consistent with the comparative requirements stated in §1.5.

5.6 Contributions and Recommendations

This study presents a lexical, phonological, morphophonological, and morphological reconstruction of Highland East Cushitic grounded in an explicit evaluation framework. The phonological component reconstructs a PHEC segmental inventory and prosodic profile on the basis of regular correspondences across the HEC languages. The morphophonological component shows how morpheme boundary alternations provide evidence for constraints and repair strategies that are not always recoverable from citation forms. Chapter 4 supplies the paradigmatic environments that make many alternations interpretable and keeps phonological generalizations grounded in attested morphological structure rather than in isolated word forms.

A central research output is the East Cushitic Lexical Database (§1.3), developed as part of this dissertation. The database is publicly available as an Excel database and an interactive web application, and it links dissertation IDs to individual lexical entries so that reconstructions, cognate sets, and supporting sources can be traced directly. Its scope is explicitly comparative across East Cushitic: It includes 23 East Cushitic languages, 356 PHEC and 375 PNHEC lexical reconstructions, in addition to reconstructed PWR parallels, and brings together over 8,500 compiled lexical entries from more than 100 scholarly sources and limited amounts of (my own) data collected (between 2024–2025) from speakers of Dahalo, Arbore, and Boni. Methodologically, it encodes reconstruction status as working hypotheses and makes uncertainty explicit by assigning each reconstructed form a Recon.Eval. confidence score (0–4), tied to correspondence regularity, semantic uniformity, and complications such as irregular correspondences or limited attestation. The EC and PWR parallels are only proposed as potential cognates.

One limitation concerns semantic and functional reconstruction. The present study is primarily a reconstruction of forms, correspondences, and distributional patterns – it hardly attempts a systematic reconstruction of semantic change criteria or a comparative reconstruction of the full range of functions associated with individual morphological exponents. These domains are nevertheless important for strengthening reconstructions and for distinguishing inherited structure from later reanalysis. In practice, semantic uncertainty is kept visible rather than normalized away: The evaluation

framework takes into account semantic stability, and glossing in examples is kept conservative (simplified glosses are used where items have multiple meanings, with the full range of attested glosses available in the database; database IDs in square brackets link examples to entries). A fuller semantics- and function-centered reconstruction would require richer and more uniform documentation than is presently available across the comparative sources assembled here.

A further limitation is uneven evidence. Some processes are widely productive but are not consistently realized across cognate sets, which restricts how far particular surface segments (especially epenthetic consonants) can be projected to the proto-level. Similarly, gaps in attestation and differences in descriptive coverage mean that negative evidence cannot be treated as decisive.

Several broader limitations also follow from the infancy of HEC comparative studies and from the design choices required by reconstruction. For some domains, the available evidence supports multiple competing hypotheses; adopting one hypothesis is sometimes necessary for consistency with dependent analyses, even where the comparative signal remains incomplete. In addition, the present study's bottom-up focus on HEC necessarily limits the degree to which wider East Cushitic comparison can be used to judge marginal correspondence patterns. Where wider comparison is likely to be decisive, the conclusions are framed conservatively.

In addition to addressing the above limitations, future work can strengthen the reconstruction in three ways. First, expanding and normalizing comparable datasets across languages would make it easier to test whether marginal correspondences reflect systematic conditioning or later remodeling. Second, deeper comparison beyond HEC, where relevant, can help distinguish inherited features from later areal convergence, especially in domains where contact is pervasive. Third, more detailed work on paradigms and morphosyntactic conditioning of terminal vowels and accent would refine the diachronic link between word-edge erosion and prosodic reorganization.

5.7 Conclusion

This study shows that reconstructing the Highland East Cushitic lexicon and phonology is most successful when the comparative method is paired with careful attention to paradigms and boundary alternations. The resulting system is typologically coherent: a constrained phonotactic profile maintained

by a limited set of repair processes, a prosodic pattern tied to syllable weight and sensitive to terminal-vowel change, and a set of morpheme boundary alternation processes that expose structure where surface forms are least stable.

The comparative database published with this research provides a resource for Cushitic linguistics and a foundation for further inquiry into Afroasiatic historical linguistics, particularly by making comparative evidence and reconstruction evaluation explicit and traceable through stable identifiers.

References

Alemu Banta. (2016). *Kookaata: Kambaatissa-Amaarsa-Ingiliizissa laaga doonnuta* [Kambaata-Amharic-English Dictionary]. Addis Ababa: Berhanena Selam Printing.

Ali, M. & Zaborski, A. (1990). *Handbook of the Oromo Language*. Prace Komisji Orientalistyczney 21. Wroclaw: Polska Akademia Nauk.

Anbessa Teferra. (2000). *A Grammar of Sidaama*. (Unpublished PhD dissertation). Hebrew University of Jerusalem.

Anbessa Teferra. (2014). *Sidaama (Sidaamu Afoo)*. Languages of the World/Materials, 501. München: LINCOM.

Andrzejewski, B. W. (1960). The Categories of Number in Noun Forms in the Borana Dialect of Galla. *Africa*, 30(1), 62–75.

Arvanites, L. (1991). *The Glottalic Phonemes of Proto Eastern Cushitic* (PhD dissertation). University of California, Los Angeles.

Bender, M. L. (1971). The Languages of Ethiopia: A New Lexicostatistic Classification and Some Problems of Diffusion. *Anthropological Linguistics*, 13(5), 165–288.

Bender, M. L. (1988). Proto-Omotic Phonology and Lexicon. In M. Bechhaus-Gerst & F. Serzisko (Eds.), *Cushitic-Omotic: Papers from the International Symposium on Cushitic and Omotic Languages*, 121–159. Hamburg: Buske.

Bender, M. L. (2020). *Cushitic Lexicon and Phonology*. Research in African Studies 28. Berlin: Peter Lang.

Bender, M. L., Bowen, J. D., Cooper, R. L., & Ferguson, C. A. (Eds.) (1976). *Language in Ethiopia*. London: Oxford University Press.

Black, P. (1974). *Lowland East Cushitic: Subgrouping and Reconstruction*. (PhD dissertation). Yale University.

Blažek, V. & Malášková, Z. (2016). Cushitic and Omotic Personal Pronouns in Afroasiatic Perspective. *Paper presented at 46th Colloquium on African Languages and Linguistics (CALL)*, Leiden University.

Corbett, G. G. (2000). *Number*. Cambridge: Cambridge University Press.

Crass, J. (2002). Ejectives and Pharyngeal Fricatives: Two Features of the Ethiopian Language Area. In Baye Yimam, R. Pankhurst, D. Chapple, Yonas Admasu, A. Pankhurst & Birhanu Teferra (Eds.), *Proceedings of the 14th International Conference of Ethiopian Studies, 6–11 November 2000,* 1679–1691. Addis Ababa: Institute of Ethiopian Studies.

Crass, J. (2005). *Das K'abeena: Deskriptive Grammatik einer hochlandostkuschitischen Sprache*. Cushitic Language Studies 23. Köln: Köppe.

Crass, J. (2007). Highland East Cushitic. In S. Uhlig (Ed.), *Encyclopaedia Aethiopica*, 3. Wiesbaden: Harrassowitz Verlag.

Crass, J. (2013). Some Remarks on the Compound Suffix Conjugation in Highland East Cushitic languages. In M.-C. Simeone-Senelle, and M. Vanhove (Eds.). *Proceedings of the 5th International Conference on Cushitic and Omotic Languages, Paris, 16–18 April 2008*, 3–20. Köln: Köppe.

Dawit Tilahun (2019). *A Grammar of Gede'o: A Highland East Cushitic Language*. (Unpublished PhD dissertation). Addis Ababa University.

Dereje Adane (2022). *Information Structure in Hadiyya*. (Unpublished PhD dissertation). Addis Ababa University.

Dimmendaal, G. J. (1983). *The Turkana Language*. Publications in African Languages and Linguistics. Dordrecht: Foris Publications.

Ehret, C. (1987). Proto-Cushitic Reconstruction. *Sprache und Geschichte in Afrika*, 8, 7–180.

Ehret, C. (1991). The Consonant Inventory of Proto-Eastern Cushitic. *Studies in African Linguistics*, 22(3), 211–276.

Ehret, C. (1995). *Reconstructing Proto-Afroasiatic (Proto-Afrasian): Vowels, Tone, Consonants, and Vocabulary*. Linguistics 126. Berkeley: University of California Press.

Esayas Tajebe (2015). *Descriptive Grammar of Saaho.* (Unpublished PhD dissertation). Addis Ababa University.

Eshetu Tadesse (2010). *Sound Correspondences and Change: A comparative Study of Burji, Hadiyya, Kambaata, Sidama and Gedeo* (Unpublished MA thesis). Addis Ababa University.

Eyob Kelemework. (2015). *A Grammar of Gedeo: A Cushitic Language of Ethiopia*. (Unpublished PhD dissertation). Osmania University.

Fleming, H. C. (1976). Omotic Overview. In M. L. Bender (Ed.), *The Non-Semitic Languages of Ethiopia*, *5*, 299–323. East Lansing: Michigan State University.

Fleming, H. C. (1988). Proto-Somotic Consonant Phonemes. In M. Bechhaus-Gerst & F. Serzisko (Eds.), *Cushitic-Omotic: Papers from the International Symposium on Cushitic and Omotic Languages, Cologne, 6–9 January 1986*, 163–175. Hamburg: Buske.

Gasparini, A. (1983). *Sidamo-English Dictionary*. Bologna: EMI.

Gasparini, A. (1995). *Grammatica Gede'o*. Trieste: Università degli Studi di Trieste, Edizioni LINT.

Getahun Addo (2017). *The Bar Burjatee, Amharic and English Dictionary*.

Gragg, G. B. (1982). *Oromo Dictionary*. East Lansing: African Studies Center, Michigan State University.

Green, C. (2021). *Somali Grammar*. Mouton-CASL Grammar Series 5. Berlin, Boston: De Gruyter Mouton.

Harvey, A. (2018). *The Gorwaa Noun: Toward a Description of the Gorwaa Language*. (PhD dissertation). SOAS.

Hayward, R. J. (1981). Nominal Suffixes in Dirayta (Gidole). *Bulletin of the School of Oriental and African Studies*, 44(1), 126–144.

Hayward, R. J. (1988). Remarks on Omotic Sibilants. In M. Bechhaus-Gerst & F. Serzisko (Eds.), *Cushitic-Omotic: Papers from the International Symposium on Cushitic and Omotic Languages, Cologne, 6–9 January 1986*, 263–299. Hamburg: Buske.

Hayward, R. J. (1984). A Reconstruction of Some Root Extensions of the Eastern Cushitic Verb. In J. Bynon (Ed.), *Current Progress in Afro-Asiatic Linguistics: Papers of the Third International Hamito-Semitic Congress, London, 1978*, 69–109. Amsterdam: John Benjamins.

Hayward, R. J. (1997). Book reviews. The Highland East Cushitic Dictionary. *Journal of African Languages and Linguistics*, 18(1), 106–120. DOI: 10.1515/jall.1997.18.1.95

Hellenthal, A-C. (2010). *A Grammar of Sheko*. Utrecht: LOT.

Hetzron, R. (1976). The Agaw Languages. *Afroasiatic Linguistics*, 3(3), 31–37.

Hetzron, R. (1980). The Limits of Cushitic. *Sprache und Geschichte in Afrika*, 2, 7-126.

Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic Explanations for the Development of Tones. *Language*, 55(1), 37–58. DOI: https://doi.org/10.2307/412518

Hudson, G. (1976). Highland East Cushitic. In M. L. Bender (Ed.), *The Non-Semitic Languages of Ethiopia*, 5, 232–277. East Lansing: Michigan State University.

Hudson, G. (1981). The Highland East Cushitic Family Vine. *Sprache und Geschichte in Afrika*, 3, 97–124.

Hudson, G. (1989). *Highland East Cushitic Dictionary*. Cushitic Language Studies 7. Hamburg: Buske.

Hudson, G. (2006). Highland East Cushitic Languages. In B. Keith (Ed.), *Encyclopedia of Language & Linguistics*, 2nd Edition, 294–298. Amsterdam: Elsevier. DOI: 10.1016/B0-08-044854-2/02073-3

Hudson, G. (2007). Highland East Cushitic Morphology. In A. S. Kaye (Ed.), *Morphologies of Asia and Africa*, 1, 529–545. Winona Lake: Eisenbrauns.

Indrias Xa'miso, Sileshi Workineh, Yohannis Latamo, & Desalegn Garsamo. (2007). *Sidama-Amharic-English Dictionary*. Addis Ababa: Master Printing Press.

Kawachi, K. (2007). *A Grammar of Sidaama (Sidamo), a Cushitic Language of Ethiopia*. (Unpublished PhD Dissertation). University at Buffalo.

Kießling, R. (1994). *Eine Grammatik des Burunge*. Hamburg: Research and Progress.

Kießling, R., & Mous, M. (2003). *The Lexical Reconstruction of West-Rift Southern Cushitic*. Cushitic Language Studies 21. Köln: Köppe.

Kramer, R., & Anbessa Teferra. (2020). Gender Switch in Sidaama. *Brill's Journal of Afroasiatic Languages and Linguistics*, 12(2), 286–327.

Lamberti, M., & Sottile, R. (1997). *The Wolaytta Language*. Köln: Köppe.

Leslau, W. (1976). *Concise Amharic Dictionary: Amharic-English, English-Amharic*. Berkeley: University of California.

Leslau, W. (1979). *Etymological Dictionary of Gurage (Ethiopic), Etymological Section*, 3. Wiesbaden: Otto Harrassowitz.

Lloret, M.-R. (1988*). Gemination and Vowel Length in Oromo Morphophonology*. (PhD dissertation). Indiana University.

Moreno, M. M. (1940). *Manuale di Sidamo*. Roma: Mondadori.

Mous, M. (1993). *A Grammar of Iraqw*. Hamburg: Helmut Buske.

Mous, M. (2016). *Alagwa: A South Cushitic Language of Tanzania: Grammar, Texts and Lexicon*. Köln: Rüdiger Köppe.

Nurse, D. (1985). Dentality, Areal Features, and Phonological Change in Northeastern Bantu. *Studies in African Linguistics*, 16(3), 243–279.

Ongaye Oda (2013). *A Grammar of Konso*. (PhD dissertation). Utrecht: LOT.

Petrollino, S. (2016). *A Grammar of Hamar: A South Omotic Language of Ethiopia*. (PhD dissertation). Leiden University.

Plazikowsky-Brauner, H. (1957). Die Hilfselemente der Konjugation in den kuschitischen Sprachen. *Zeitschrift Der Deutschen Morgenländischen Gesellschaft*, 107 (n.F. 32)(1), 7–30. Retrieved from http://www.jstor.org/stable/43374561

Ritter, S. (2007). *Hadiyya-English-Dictionary* (Unpublished manuscript).

Roba Dame & Wedekind, C. (2008). *Burji Dictionary* (Unpublished manuscript).

Sasse, H.-J. (1976). Weiteres zu den ostkuschitischen Sibilanten. *Afrika und Übersee: Sprachen, Kulturen*, *59(2)*, 125–142.

Sasse, H.-J. (1979). The Consonant Phonemes of Proto-East-Cushitic (PEC): A First Approximation. *Afroasiatic Linguistics*, 7(1), 1–67.

Sasse, H.-J. (1982). *An Etymological Dictionary of Burji*. Hamburg: Buske.

Schneider-Blum, G. (2007). *A Grammar of Alaaba: A Highland East Cushitic Language of Ethiopia*. Cushitic Language Studies 25. Köln: Köppe.

Sim, R. J. (1991). Hadiyya's Other Causative? In H. G. Mukarovsky (Ed.) *Proceedings of the Fifth International Hamito-Semitic Congress: Cushitic, Egyptian, Omotic, Semitic, Vienna, 27th September - 2nd October 1987*, 49–64. Wien: Afro-Pub.

Stroomer, H. (1987). *A Comparative Study of Three Oromo Dialects in Kenya: Phonology, Morphology and Vocabulary*. Cushitic language studies 6. Hamburg: Helmut Buske Verlag.

Stroomer, H. (2001). *A Concise Vocabulary of Orma Oromo (Kenya): Orma-English, English-Orma*. Cushitic Language Studies 16. Köln: Köppe.

Tadesse Sibamo. (2015). *Documentation and Description of Hadiyya (A Highland East Cushitic Language of Ethiopia)*. (Unpublished PhD dissertation). Addis Ababa University.

Tesfaye Baye. (2015). *A Descriptive Grammar of Burji*. (Unpublished PhD dissertation). Addis Ababa University.

Tilahun Gamta. (2004). *Galmee Afaan Oromo fi Afaan Ingilizii barasiisuu = Comprehensive Oromo-English Dictionary (COED)*. New York: Karrayyuu Publishing.

Tosco, M. (1991). *A Grammatical Sketch of Dahalo: Including Texts and a Glossary*. Cushitic Language Studies 8. Hamburg: Buske.

Tosco, M. (1994). The Historical Syntax of East Cushitic: A First Sketch. In T. Bearth, W. J. G. Möhlig, B. Sottas, & E. Suter (Eds.), *Perspektiven Afrikanistischer Forschung. Beiträge Zur Linguistik, Ethnologie, Geschichte, Philosophie Und Literatur / X. Afrikanistentag*, *Zürich, 23rd–25th September 1993*, 415–440. Köln: Rüdiger Köppe.

Tosco, M. (2020). East Cushitic. In R. Vossen and G. J. Dimmendaal (Eds.), *The Oxford Handbook of African Languages*, 290–299. London: Oxford University Press.

Treis, Y. (2008). *A Grammar of Kambaata. Part 1: Phonology, Nominal Morphology, and Non-Verbal Predication*. Cushitic Language Studies 26. Köln: Köppe.

Treis, Y. (2022). Verbal Derivation in Kambaata (Cushitic), with a Focus on the Encoding of Noncausal/Causal Alternations. *Linguistique et langues africaines*, 8(2). DOI: 10.4000/lla.4535

Treis, Y. (2024). Ideophones in Kambaata (Cushitic): Grammar, Meaning and Use. *Linguistic Typology at the Crossroads*, 4(1), 147–200.

Treis, Y. (p.c.). Kambaata Wordlist (Unpublished fieldwork data).

Wedekind, K. (1980). Sidamo, Gedeo (Derasa), Burji: Phonological Differences and Likenesses. *Journal of Ethiopian Studies*, 14, 131–176.

Wedekind, K. (1985). Gedeo (Derasa) Verb Morphology and Morphophonologicals. In Abebe Gebre-Tsadik, M. Sim, R. J. Sim, C.

Wedekind & K. Wedekind (Eds.), *The Verb Morphophonologicals of Five Highland East Cushitic languages including Burji*, 82–109. Afrikanistische Arbeitspapiere 2. Köln: Cologne University.

Wedekind, K. (1990). *Generating Narratives: Interrelations of Knowledge, Text Variants and Cushitic Focus Strategies*. Trends in Linguistics: Studies and Monographs, 52. Berlin: Mouton de Gruyter.

Wedekind, K. (2008). *Gedeo Dictionary* (Unpublished manuscript).

Wondwosen Tesfaye (2020). Gender and Number in Mosittacha. *ZENA-LISSAN (Journal of Academy of Ethiopian Languages and Cultures*), 29(1), 103–130. Retrieved from http://213.55.95.79/index.php/JAELC/article/view/1769

World Bank. (2026). *Health Nutrition and Population Statistics: Population estimates and projections* [Data set]. DataBank. Retrieved from https://databank.worldbank.org/reports.aspx?source=Health%20Nutrition%20and%20Population%20Statistics%3A%20Population%20estimates%20and%20projections

Yri, K. M. (2004). Orthography and Phonology in Sidaamu Afoo (Sidamo). *Journal of Ethiopian Studies*, 37(1), 41–56.

Yri, K. M., & Pepper, S. (2019). *Sidaama Dictionary*. Dictionaria, 6, 1-3578. DOI: 10.5281/zenodo.5526518

Zaborski, A. (1986). *The Morphology of the Nominal Plural in the Cushitic Languages*. Beiträge zur Afrikanistik 28. Wien: Institute für Afrikanistik und Ägyptologie der Universität Wien.

Appendix

This appendix presents the reconstructed lexicon referenced in the study. The entries are listed by their unique database identifier (ID) to facilitate cross-reference with the East Cushitic Lexical Database. For each item, the Proto-Highland East Cushitic (PHEC) and Proto-North Highland East Cushitic (PNHEC) reconstructions are provided where available, along with their glosses.

ID	PHEC	PNHEC	Gloss
[1]	***il-**	***ill-**	‘eye’
[2]	***an-i**	***ani**	‘I’
[3]	***aabb-**	***aabb-**	‘father’
[4]	***gulib-, *gilub-**	***gulub-**	‘knee’ (see [430])
[5]	***san-**	***san-**	‘nose’
[6]	***is-i**	***isi**	‘she’
[7]	***is-u**	***is-u**	‘he’
[8]	***ʃool-**	***ʃool-**	‘four’
[9]	***arrab-**	***arrab-**	‘tongue’ (see [548])
[10]	***hat’t’iʃ y**	***hat’t’iʃ- y-**	‘to sneeze’
[11]	***is-n-**	***is-n-**	‘they’
[12]	***maħ-**	***maħ-**	‘what?’
[13]	***an-ʃ-**	***an-ʃ-**	‘saliva’ (see [807])
[14]	***az-**	***az-**	‘milk’
[15]	***bezz-k-**	**-**	‘star(s)’
[18]	***t’iig-**	**-**	‘blood’
[19]	***sazhet**	***sazzeett-**	‘eight’
[20]	***sazeħ-**	***saseħ-**	‘three’
[21]	**-**	***c’ak’as-**	‘to listen, spy’
[22]	***ħanʔur-**	***ħaʔur-**	‘navel’
[23]	***hiir-**	**-**	‘to untie’
[27]	***at-i**	***ati**	‘you’ (sg.subj)
[28]	***ay- (→ *ay-y-e/*ay-y-i)**	***ay- (→ *ay-y-e/*ay-y-i)**	‘who?’ (obj, subj)
[29]	***k’ab-**	***ab-**	‘to hold, seize’
[30]	***kaʔ-**	***kaʔ-, *keʔ-**	‘to get up, arise’
[31]	***oo y, *oo ʔ**	***oo y, *oo ʔ**	‘to cry’ (see [235])
[32]	***ʔ-ʔ-**	***ʔ-ʔ-**	‘to take, receive’
[33]	**-**	***osal-ʔ-**	‘to laugh’
[34]	***baʔ-**	***baʔ-**	‘be lost’ (see [111])
[36]	***gam-**	***gam-ʔ- → *gaʔm-**	‘to bite’
[38]	***liħ-**	***leħ-**	‘six’

[39]	***suk-ʔ-**	***suk'k'-**	'to spin'
[40]	***tuf-**	***tuf-**	'to spit'
[42]	**-**	***k'eer-**	'long, tall'
[44]	**-**	***c'aʔ-** → ***c'aʔ ʔ, *c'aʔ y, *c'aʔ-i-ʃʃ-**	'to clean'
[45]	***fiik' ʔ**	***fiik' ʔ**	'to whistle'
[46]	***ik'k'-, *ikk-**	***ink'-, *ink-**	'tooth'
[47]	***k'uf-**	***kuf-** (→ ***kuf-t-ʔ-** / ***kuf-s-ʔ-**)	'to cough'
[48]	***tum-**	***tum-**	'to pound' (see [254])
[49]	***buk'-buk'-**	***k'uub-** (→ ***k'uub-ʔ-**)	'egg'
[51]	***lokk-, *lekk-**	***lokk-, *lekk-**	'foot, leg'
[52]	**-**	***hit'-, *ays-/*ans-**	'grass (sp.)'
[54]	***min-**	***min-**	'house'
[55]	***t'ut'-**	***t'ut'-**	'black ant'
[59]	***agan-**	***agan-**	'month'
[62]	***al-**	***il-**	'give birth, beget (of human)' (see [449])
[63]	***naʔ-**	***naʔ-**	'we'
[65]	**-**	***faras-, *faras-ʃ-**	'horse'
[66]	***anuun-**	***anuun-**	'breast'
[67]	***ang-**	***ang-**	'hand, arm'
[68]	***malaħ-**	***malaħ-**	'pus'
[69]	***mizk-**	***mizk-**	'right(-side)'
[70]	***afoʔ-**	***afoʔ-**	'mouth, opening, language'
[71]	***k'utaħ-**	***utaħ-**	'thorn'
[72]	***lam-**	***lam-**	'two'
[74]	***k'ub-, *-kub-**	***ub-**	'to fall (down)'
[75]	***t'eʔm-**	***t'eʔm-**	'to taste good'
[79]	***eey-**	***ee, *yee**	'yes'
[80]	***gal**	***gal-**	'to enter'
[81]	**-**	***kob-ʔ-, *c'amm-**	'shoe, sandal'
[84]	***c'ab-**	***c'abb-**	'to cut, sharpen'
[85]	***arg-**	***arg**	'to send'
[86]	***is-n-**	***is-n-** → ***ins-**	'them'
[87]	***k'ot-**	***k'ot-**	'to dig'
[88]	***gudum-**	**-**	'shoulder'
[90]	**-**	***malab-**	'honey'

[91]	***mooyy-**	***mooyy-, *udulun-m-a**	‘mortar’
[94]	***gorɑʔ-**	**-**	‘to slaughter (cattle)’
[95]	**ʃomb-**	**-**	‘lung’
[103]	***e-y**	***e-y**	‘me’
[104]	***ati-n-**	***aʔ-n-**	‘you (pl.subj)’
[106]	***t’aban-**	***t’aban-**	‘side, cheek’ (see [214])
[108]	***sunk’**	***sunk’**	‘to kiss’
[109]	***ʃ- → *ʃ y**	***ʃ- → *ʃ y**	‘to kill’
[110]	***bar-**	***barr-**	‘day’ (see [361])
[111]	***baʔ-is-**	***baʔ-is-**	‘to spoil, extinguish’ (see [34])
[112]	***bubk-**	***bub-**	‘burn’
[113]	**-**	***dabar-**	‘to add, to answer’
[114]	***debeʔ-**	***debeʔ-**	‘to belch’
[115]	***k’ab y-t-**	***k’iid-**	‘become cold’
[116]	***k’ok’k’-**	***k’oonk’-**	‘throat’
[117]	***saʔ-**	***saʔ-**	‘cow’
[118]	***zaak-**	***zaaz-, *zaak**	‘to swim, to spil’
[120]	***fil-ʔ-**	***fil-ʔ-, *fil-t-, *fil-t-ʔ-**	‘to comb’
[121]	***gol-**	***gol-**	‘wall’
[122]	***gor-**	***gor-**	‘thorn bush’
[125]	***buk-**	***buk-**	‘dust’
[126]	***gog-**	***gog-**	‘skin, hide’
[127]	**-**	***kontor-**	‘to snore’
[128]	***t’eeb ʔ, *t’eeb-**	***t’eeb ʔ, *t’eeb-**	‘to be thirsty’, ‘thirst’
[130]	***tuul-**	***tuul-, *tuur-m-**	‘to pile up’
[131]	**soog-dd**[213]	**-**	‘salt’
[135]	***ʔ-y**	***ʔ-y**	‘my’
[136]	***eel-**	***eel-**	‘pond, pool’
[137]	***lal-**	***lal-**	‘cattle’
[138]	***man-ʃ-**	***man-ʃ-**	‘person’ (see [215], [609])
[139]	***wozan-**	***wozan-**	‘heart’
[140]	**-**	***ind-i-ʔind-**	‘tears’
[141]	***k’ar-**	***k’ar-**	‘to sharpen’
[142]	***tum-am-ʃ-, *tum-t-, *tum-t-iʃʃ-**	***tum-am-ʃ-, *tum-t-, *tum-t-iʃʃ-**	‘ten’ (see [254])
[143]	***fur-**	***fur-**	‘to open’

[213] Form listed without an asterik are just for reference but they are not reconstructed to P(N)HEC).

[144]	***meʔ-**	***meʔ-**	‘how many? how much?’
[145]	***sum-, *sim-**	***ʃum-**	‘urine’ (see [170])
[150]	***baall-**	***baall-**	‘feather’ (see [276])
[151]	***is-e**	***ise**	‘her’
[161]	***bark-**	***bark-**	‘head-support, pillow (of wood)’
[162]	***fic’c’- ← fiʃ-ʔ-, *fil-s-ʔ-**	***fi-n-c’-**	‘to scatter’
[163]	***ħur-ʔ-**	***ħur-**	‘young ox, bull’
[164]	***t’uf-**	***t’uf-**	‘to close’
[167]	***fayy-**	***fayy-**	‘healthy’
[168]	***heeʔ-, *heeʔ-ʃ-**	***heeʔ-, *heeʔ-ʃ-**	‘to exist, to live, life, existence’
[169]	***k’unc’-**	**-**	‘bark (of tree)’
[170]	***sum ʔ**	***ʃum ʔ**	‘urinate’ (see [145])
[173]	***billaa-t**	***billaa-t**	‘butterfly’
[174]	***hamas-**	***hamas-, *hamas-ʃ-**	‘snake’
[175]	***kalt-**	**-**	‘axe, hoe’
[180]	***utub-**	***utub-**	‘pole, center-pole of house’
[183]	***gib-**	***gib-**	‘to refuse, hate’
[184]	***reħ ʔ**	***reħ ʔ**	‘die’ (see [329], [439])
[186]	***c’ub-**	***c’ub- → *c’ub ʔ, *c’ub y-s**	‘dip in, immerse’ (see [447])
[187]	***kin-ʔ-**	***kiʔn-**	‘your’
[188]	***ni-ʔ-**	***ni-ʔ-**	‘our’
[189]	***k’araar-**	***k’araar-**	‘be bitter’
[190]	**-**	***arr-, *arr-i-ʃʃ-**	‘sun’
[191]	**-**	***ull-**	‘earth’
[195]	***fal-, *falfal-**	***fal-**	‘omen, magic’
[196]	***ki-n-**	***kin-ʔ- → *kiʔn-e, *kiʔn-o**	‘you’ (pl)
[197]	***n-y**	***n-y**	‘us’
[198]	***tum-am-ʃ-, *tumt-i-ʃʃ-, *tumt-**	***tum-am-ʃ-, *tumt-i-ʃʃ-, *tumt-**	‘blacksmith’ (see [142], [254])
[199]	***ul-**	**-**	‘to smell’
[202]	***t’all-**	***t’all-**	‘only, alone’
[206]	***awd-**	***awd-**	‘threshing floor’
[207]	***ħun-, *ħut-**	***ħun-, *ħut-**	‘strength, power’
[208]	***ħur ab-**	***ħur-**	‘to pour’
[209]	***laaf-**	***laaf-**	‘soft, weak’
[212]	**-**	***ar- → *ar-o**	‘husband’ (see [271])

[214]	***t'aban ʔ → *t'aban ʔ-s-**	***t'aban ʔ-s-**	'to slap' (see [106])
[215]	***meen-t-**	***man-t-, *meen-t-**	'woman, girl' (see [138], [609])
[217]	**-**	***fool-**	'breath, odor, smell'
[220]	***c'ark'-**	***c'ark'-**	'rags, cloth'
[221]	***c'inc'-**	**-**	'good smell'
[222]	***t'ok-**	***t'ook-**	'to burst'
[223]	***eg-**	***eger**	'to wait'
[224]	***t'ur-**	***t'ur-**	'dirt of body, clothing'
[226]	***c'iwʔ-, *c'awʔ-**	***c'aac'urr-, *c'iic'oorr-**	'chick(en)' (see [348])
[228]	***harr-**	***harr-, *harr-i-ʃʃ-**	'donkey'
[229]	***haʃ~haʃ-**	***haʃ~haʃ-**	'to whisper' (see [252])
[235]	***od-**	**-**	'news' (see [31])
[236]	***og-, *og ʔ-y-ʃ-**	***og-, *og ʔ-y-ʃ-**	'traditional doctor'
[237]	***ʃirb**	***sirb-**	'to dance and sing'
[240]	***k'ort'umm-**	***k'ort'umm-**	'fish'
[249]	**-**	***ʃik' y, *ʃik' ʔ**	'to approach'
[250]	***t'iib-**	***t'iib-**	'to push'
[252]	***hasaaw-, *haas-**	***hasaaw-, *haas-**	'speak, talk, speech' (see [229])
[253]	***ki-ʔ-**	***ki-ʔ, *-ʔ-ki**	'your'
[254]	**-**	***tun~tumm- ← *tum~tum-**	'fist' (see [142])
[256]	***ham ʔ**	***ham ʔ**	'to slander'
[257]	***ke**	***ke**	'you' (sg.obj)
[258]	***maak-, *ma~mmaak-**	***maak-**	'say in a proverb, tell a story'
[259]	***wol-**	***wol-**	'(each) other'
[264]	***arr-**	**-**	'grey hair'
[266]	***ħur ʔ**	***ħurk- → *ħurk-ʔ, *ħur-ʃ-ʔ**	'to choke'
[267]	***t'ulak-**	***t'ulak-**	'claw, fingernail'
[268]	***mur-**	***mur-**	'to cut'
[269]	***t'aʔ-, *t'eʔ-**	***t'aʔ-, *t'eʔ-**	'to extinguish, to diminish'
[271]	**-**	***ar- → ar-e**	'wife' (see [212])
[273]	***biin-**	***biin-**	'mosquito, fly (sp.)'
[274]	***yeyy-**	**-**	'wolf, jackal'
[276]	***ball-**	***ball-**	'brother-in-law, father-in-law'
[279]	***isa-ʔ-**	***isiʔ-**	'his'
[281]	***k-, *t-**	***ku**	'these' (m, f)
[282]	***aar-**	**-**	'be angry'

[284]	***ayyaan-**	***ayyaan-**	‘festival, holiday’
[285]	***balʔ-**	***bal-ʔ-**	‘broad, wide’
[286]	***c’ob-**	***c’ob-ʔ-** → ***c’op’p’ y**	‘to drip, to leak’
[287]	***t’ib- (ʔ)**	***t’ib-**	‘to make trouble’ (see [417])
[288]	***k’el-**	***k’el-**	‘snail’
[292]	***is-o**	***is-o**	‘him’
[293]	***waʔ-**	***waʔ-**	‘water’
[298]	***al ʔ**	***el ʔ**	‘to be useful’
[299]	**-**	***zaane-i-ʃʃ-**	‘elephant’
[300]	***k’ub-**	***k’ub-**	‘finger(-ring), ring’ (see [494])
[304]	***wit’-**	***wit’-**	‘sow, to scatter’
[306]	***y- ebb-**	***y- ebb-**	‘be(come) hot, warm’
[307]	***mit-**	***mat-, *mit-**	‘one’
[308]	***t’uur-**	**-**	‘to milk’ (see [351])
[309]	***war-**	***war-**	‘to gossip, tell the news’
[316]	***dam-**	***dam-**	‘branch, stick’
[318]	***isn-ʔ-**	***isn-ʔ-**	‘their’ (poss)
[321]	***bukk-**	***bukk-, *bukk-ʔ-**	‘to pluck’
[325]	***duud-**	***duud-**	‘mute (person)’
[326]	***geeg-ess-**	***geeg-eess-**	‘accompany’
[327]	***kor-**	***koor-**	‘saddle’
[328]	***k’ob-**	***k’ub-**	‘moisture, dew’
[329]	***reħ-ʃ-**	***reħ-ʃ-**	‘corpse’ (see [184], [439])
[330]	***sooz-**	***sooz-**	‘dawn’
[338]	**-**	***yaʔ-**	‘to call’ (see [749], [354])
[339]	***ful-**	***ful-**	‘to go out, exit’
[340]	***gag~gal-** → ***ga-n-gal-**	***gog~gol-** → ***go-n-gol-**	‘to roll’
[341]	***loom-**	***loom-**	‘lemon’
[345]	**-**	***raar-**	‘shout, scream’
[347]	**-**	***oos-**	‘boy’
[348]	***c’iiʔ-**	***c’iiʔ-i-ʃʃ-**	‘bird’ (see [226])
[349]	***gabal-**	***gabal-**	‘border, limit’ (loanword)
[350]	***sang-**	***sang-**	‘ox, bull’
[351]	***t’uut’-**	***t’uut’-**	‘to suck, absorb’ (see [308])
[354]	***yaaʔ-**	***yaaʔ-**	‘to flow, call, assemble’ (see [749], [338])
[361]	***bar-**	***ber-**	‘yesterday’ (see [110])
[362]	***moor ʔ**	***moor ʔ**	‘to steal’ (see [708])

[366]	***ise-ʔ-**	***ise-ʔ-**	‘her (poss)’
[367]	***mic’ir**	**-**	‘twist, squeeze’
[377]	***c’anc’-**	***c’anc’-**	‘to make noise’
[378]	***dan-**	***dan-t-**	‘to be able’
[380]	***hik’k’-, *hek’k’-**	***hik’k’-, *hek’k’-**	‘have hiccups’
[381]	***k’ul-, *k’ul~k’ul~l-ʔ-**	***ul-, *ul~ul~l-ʔ-**	‘leech’
[382]	***oll-**	***oll-**	‘neighbour’
[383]	***orz-**	***orz-, *orza-**	‘hair of animals’
[388]	***y → *is- (← *y-s-**)	***y → *is- (← *y-s-**)	‘to say, do, be’ (see [425])
[389]	***duk-**	***duk-**	‘to sneak, creep’
[390]	***hiit’-**	***hiit’-**	‘to stretch’
[391]	***kum-**	***kum-**	‘thousand’
[392]	***lik’-**	***lik’-**	‘to lend’
[397]	**-**	***kikkil-**	‘armpit’
[399]	***dawm-**	***dawam-**	‘red’
[400]	***giir-**	***giir-**	‘fire’
[406]	**-**	***wees-**	‘ensete tree (Ensete ventricosum)’
[407]	***ebel-, *ebal-**	***ebel-, *ebal-**	‘so-and-so’
[408]	***mut-**	***mut-**	‘awl for basketwork’
[411]	**-**	***buur-**	‘butter’ (see [742])
[415]	***afal-**	***afal-**	‘liver’
[416]	**-**	***c’in- → *c’in-k, *c’in-s, *c’in ʔ**	‘to squeeze’
[417]	***t’ib-**	***t’ib-**	‘disease, evil’ (see [287])
[418]	***ku/*ti, *ko/*te / *ku/*ti, *ka/*ta**	***ku/*ti, *ko/*te**	‘this’ (m/f)
[425]	***ʔ, *ass-**	***ʔ, *ass- (← *ʔ-s-)**	‘to do’ (see [388])
[426]	***ħaraar y-s-, *araar y-s-**	***ħaraar y-s-, *araar y-s-**	‘to mediate, reconcile’
[428]	***dur-**	**-**	‘rich’
[429]	***geedaal-**	**-**	‘fox’
[430]	***gulib ʔ, *gilub ʔ**	***gulub ʔ**	‘to kneel’ (see [4])
[431]	***raʔ-**	***raʔ-**	‘ripen, be ripe’ (see [496])
[436]	**-**	***ad-**	‘aunt’
[437]	***seer-**	***seer-**	‘law’
[438]	***zuug-**	***zuug-**	‘to scrape, tan (hide)’
[439]	***reħ-**	***reħ-**	‘death’ (see [329],[184])
[440]	***weeʔn-**	***weeʔn-, *weeʔn-i-ʃʃ-**	‘Colobus monkey’

[443]	***uħw-**	***uħw-**	‘to give’
[444]	***bagaz-**	***bagaz-**	‘spear’
[445]	***bork-**	***bork-**	‘to be impure’
[447]	***c’ubb-**	***c’ubb-**	‘sin’ (see [186])
[448]	***gaayy-**	***gaayy-**	‘pipe (of clay)’
[449]	***k’al-**	***k’al-**	‘give birth, beget (of animals)’ (see [62])
[450]	***ʃaal-**	***saal-**	‘be shy, be ashamed’
[460]	**-**	***beet-**	‘son, child’ (see [741])
[461]	***has-**	***has-**	‘to want, seek’
[462]	***meet-**	***meet-**	‘scraping board for use in preparing ensete’
[466]	***diwh-, *diwk-**	***diwh-, *diwk-**	‘empty’
[467]	***ħons-**	***ħons-**	‘nine’
[476]	***c’ir-**	***c’ir-**	‘to clear’ (see [516])
[477]	***hobol-**	**-**	‘boat’
[478]	***kott-**	***kott-**	‘hoof, foot’
[490]	***c’obt’-**	***c’obt’-**	‘gonorrhea’
[491]	***fac’-**	***fac’-**	‘to cut, break’
[492]	***haar-**	***haar-**	‘new’
[493]	***hiiʔ-**	***hiiʔ-**	‘take!’
[494]	***k’ob-, *k’ok’~k’obb-**	***k’ob-, *k’ok’~k’obb-**	‘to surround, to bend’ (see [300])
[495]	***lab-**	***lob-**	‘many, big, much’
[496]	***raʔ-i-s-**	***raʔ-is-**	‘to cook’ (see [431])
[497]	***ʃakol-, *ʃol-**	***ʃoll ʔ**	‘be easy’
[511]	***zal~zal-**	***zaz~zal-**	‘trade, business’
[513]	**-**	***c’aat-**	‘chat (Catha edulis)’
[514]	***ebb-**	***ebb-**	‘to bring’
[516]	***c’il**	***c’il-**	‘to defecate’ (see [476], [516])
[517]	***c’iil-**	***c’iil-**	‘excrement’ (see [516])
[519]	***ħililll-**	***ħililll-**	‘smoke’
[520]	***ħuy- → *ħuy ʔ-m-**	***ħuy ʔ-m-**	‘be(come) full’ (see [607])
[523]	***ʃam-**	***ʃam-**	‘to rot, to be wet’
[544]	**-**	***c’ub-**	‘dagger, knife’
[545]	***dun-**	***dun-am-**	‘to leak, poke’
[547]	***am-**	***am-**	‘mother’
[548]	***arraab-**	***arraab-**	‘to lick’ (see [9])
[550]	***ibb-**	***ib-, *ib~ib-**	‘louse’
[551]	***it-**	***it-**	‘to eat’
[580]	***boc’-**	***boc’-**	‘to carve, split (wood)’
[581]	***zar-**	***zar-**	‘to tear’
[582]	***hak’k’-**	***hak’k’-**	‘wood’

[583]	***maaz ʔ, *miiz ʔ**	***maaz ʔ, *miiz ʔ**	‘to hurt’
[601]	**-**	***hos-**	‘to pass, spend the day’
[602]	**-**	***kin-**	‘stone, rock’
[603]	***boos-**	***boos-**	‘fireplace’
[604]	***c’al-**	***c’al-, *k’al-**	‘to hang’
[605]	***galat-**	***galat-**	‘thanks’
[606]	***gubeed-**	***gubeed-**	‘thigh’
[607]	***ħuy ʔ-s-**	***ħuym-s-**	‘to fill’ (see [520])
[609]	***man~n-**	***man~n-**	‘people’ (see [138], [215])
[610]	***roob-**	***roob-**	‘hippopotamus’
[613]	***t’een-**	***t’een-**	‘rain’
[623]	***oot-**	***oot-**	‘fence’
[627]	**-**	***zoħ ʔ**	‘to wander’
[628]	***an~n-**	***ann-**	‘father’
[629]	***bad-**	***bad-**	‘to mistreat’
[632]	***gidd-, *gudd-**	***gidd-, *gudd-**	‘inside’
[633]	***has y-s-**	***has-is-**	‘be necessary’
[634]	***leem-**	***leem-, *leem-n-**	‘bamboo’
[635]	***maal-**	***maal-**	‘meat’
[636]	***omt-, *om-omt-**	***ont-**	‘five’ (see [254])
[637]	***sim-**	***sim-**	‘ensete tree 1st stage’
[638]	***sor-**	***sor-**	‘red soil’
[639]	***ʃaf-, *ʃaf-ʃaf- ʔ**	***ʃaf-, ʃaf-ʃaf- ʔ**	‘to shake’
[652]	***kiil-**	***kiil-**	‘to do magic’
[660]	**-**	***k’oʔ-**	‘marrow of bones’
[661]	**-**	***t’umm-**	‘peace’
[670]	**-**	***kaas-**	‘to plant’
[671]	**-**	***wad-**	‘to shout’
[672]	***dul-**	***dul-**	‘to skin, slaughter’
[673]	***ik-**	***ik-, *ikk-**	‘to be(come), happen’
[675]	***k’oom-**	***k’oom-**	‘calabash’
[693]	**-**	***darʃ-**	‘to swell’
[698]	***ħokol-**	***ħok~kol-**	‘to roll’
[700]	***ham-ʔ-**	***haʔm-i-ʃʃ-**	‘ensete tree root/corm’
[701]	***gig-**	***gigg-is-**	‘to destroy’
[702]	***hig-**	***hig-**	‘to return, pass’
[703]	***kaakk y-s-**	***kaakk y-s-**	‘to cackle (hen)’
[704]	**-**	***kaʃil-**	‘embers, charcoal’
[707]	***magan-**	***magan-**	‘sky, God’
[708]	***moor-k-**	***moor-k-**	‘thief’ (see [362])
[721]	***fir-**	**-**	‘relative’ (see [744])
[722]	***hoog-**	***hog-**	‘to clean’
[723]	***sek- → *seh- (→ *seh-s-)**	***sek- → *seh- (→ *seh-s-)**	‘adorn, decorate’

[730]	**-**	***t'aʔm-**	'to ask, visit'
[741]	**-**	***beet-, *beet-beet-**	'grandchild' (see [460])
[742]	**-**	***buur-, *buur-ʔ-**	'to smear' (see [411])
[744]	***firt'-**	***fit't'-**	'to disperse' (see [721])
[745]	***-n**	***-n, *-nni**	'with'
[749]	***yaaʔ-**	***yaaʔ-**	'assembly, meeting' (see [354], [338])
[768]	***hab-**	***hab-**	'to forget'
[777]	**-**	***dirr-**	'to descend, go down'
[807]	***anʃ-**	***inc'-**	'to chew, twist' (see [13])
[1030]	***ħuk'-**	***ħuk'-**	'be(come) thin'
[1032]	**-**	***nigg-**	'tendon'
[1034]	***ak-**	**-**	'grandfather'‘