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).

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).

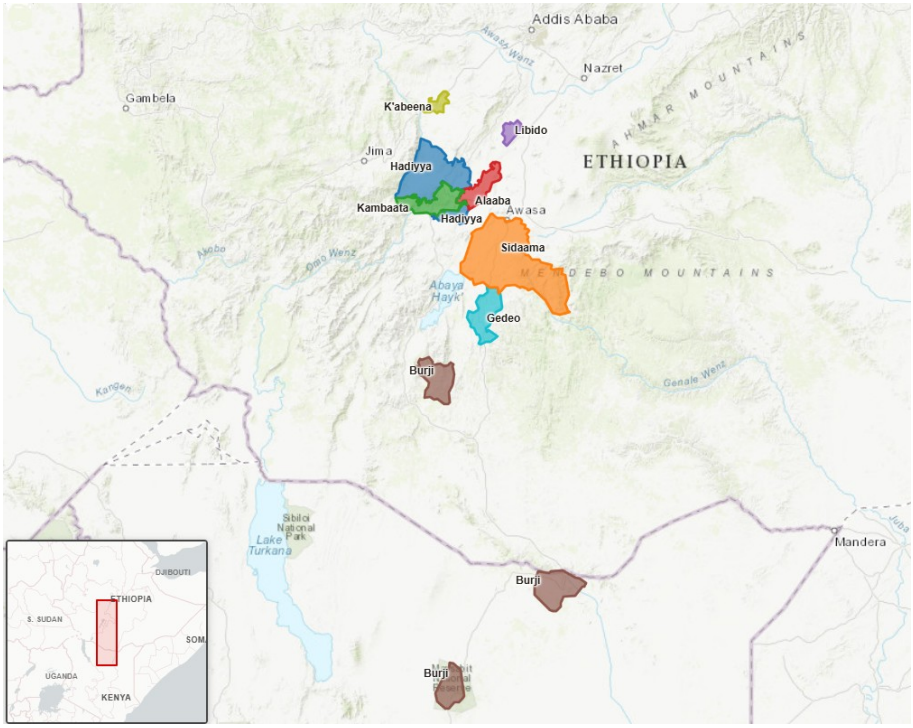

Figure 1: *Map of the Highland East Cushitic Languages*

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:

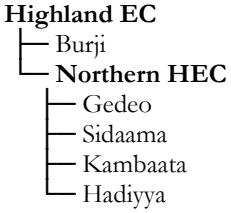

Figure 2: *Highland East Cushitic Family Tree*

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.