Multilingualism Doctoral School

Faculty of Humanities

University of Pannonia

A study on code-switching between Afaan Oromoo and English in an informal context

Ph.D. Dissertation

written by

Mihiretu Wakwoya

Supervisor:

Prof. Margaret Deuchar

Veszprém, Hungary

2026

Table of Contents

ABSTRACT

This study examines code-switching between Afaan Oromoo (Maccaa dialect) and English in informal conversations in the Kellem Wollega Zone of Western Ethiopia. It focuses primarily on the structural relationship between Afaan Oromoo and English while also systematically documenting Amharic insertions that arise naturally in the recordings of audio conversations. Three core questions guide the analysis: which language functions as the matrix language in code-switched clauses and how this shapes word order and clause structure; how English lexical items are morphosyntactically integrated into Afaan Oromoo; and how far the Matrix Language Frame (MLF) and 4-M models account for the structural patterns observed in this multilingual environment. The recordings comprise 21 hours of spontaneous conversation involving 56 (F=18, M=38) bilingual speakers. For the quantitative analysis, three recordings were fully transcribed, segmented at the clause level, and annotated using an adapted CHAT-based format combined with Leipzig glossing rules, yielding 879 analyzable clauses. In the quantitative analysis, examples were taken from other recordings that were not part of the quantitative analysis. The results show that Afaan Oromoo overwhelmingly serves as the matrix language in both monolingual and bilingual clauses, providing word order, system morphemes, and agreement, while English and Amharic function primarily as sources of content morphemes. Of the 739 monolingual clauses, 728 are in Afaan Oromoo and 11 in Amharic in the fully transcribed subset. Among the 140 bilingual clauses, Afaan Oromoo is the matrix language in virtually all cases, with only a few exceptional Amharic-matrix clauses, while English appears as an embedded language in 70 clauses and Amharic in 54, with 6 clauses containing both English and Amharic insertions. English nouns are integrated through Afaan Oromoo case marking, possessive and focus morphology, whereas English verbs are incorporated mainly via light verb constructions in which an Afaan Oromoo verb carries all inflectional material; similar integration strategies are observed for Amharic-origin items, which often occur in the same syntactic slots but are more frequent in domains tied to local administration and everyday interaction. Questionnaire data indicate that most participants acquired Afaan Oromoo as a first language, received their early schooling through Afaan Oromoo, and encountered English—and for many, Amharic—later through formal education and wider social networks. Participants express strong instrumental and affective attachment to Afaan Oromoo and generally positive attitudes toward English, yet many also claim that languages should not be mixed while simultaneously reporting that they do not in practice keep Afaan Oromoo, English, and Amharic separate in everyday conversation. The study makes four main contributions. Theoretically, it extends the empirical basis of the MLF and 4-M models to an agglutinative Cushitic language in contact not only with English but also with Amharic, demonstrating that late outsider system morphemes consistently come from the matrix language even in three-way contact settings. Methodologically, it shows how a combined CHAT–Leipzig framework and multi-tier spreadsheet implementation can be adapted for fine-grained, clause-level analysis in an under-resourced language. Empirically, it offers the first detailed structural account of Afaan Oromoo–English code-switching that also explicitly incorporates Amharic contact patterns in informal interaction. Practically, the findings suggest that Ethiopian language and education policy, as well as classroom practice, can benefit from recognizing code-switching and cross-linguistic integration as systematic manifestations of bilingual and trilingual competence, rather than deviations from an ideal of strict language separation.

Key words: Afaan Oromoo, bilingualism, code-switching, Mecha dialect, morphosyntactic integration,

CHAPTER ONE. INTRODUCTION

1. Introduction

This chapter introduces the sociolinguistic context of Ethiopia and the significance of examining code-switching (CS) between Afaan Oromoo and English in informal interaction. It outlines Ethiopia's linguistic diversity, identifies a gap in research on CS involving the Mecha (Maccaa) dialect of Afaan Oromoo, and presents the aims, research questions, hypotheses, and overall structure of the study.

1.1 Background of the Study

Ethiopia's linguistic landscape is characterized by profound diversity, with more than 80 languages spoken across its regions, reflecting a rich ethnic mosaic and a complex history of contact that has fostered bilingualism and multilingualism (Leyew, 2012; Meyer et al., 2023; Woldemariam & Lanza, 2014). These languages are commonly grouped into several major families—Afroasiatic, Nilo-Saharan, and Omotic—each comprising numerous individual languages (Ado et al., 2021; Meyer et al., 2023). The Afroasiatic family, the largest in Ethiopia, includes Amharic (the federal working language), Afaan Oromoo, Tigrinya, Somali, and Afar, which together cover large parts of the country (Ado et al., 2021; Ammon et al., 2008; Meyer et al., 2023; Woldemariam & Lanza, 2014). Nilo-Saharan languages such as Nuer, Anuak, and Majang are primarily spoken in western and southwestern areas, while Omotic languages such as Wolaytta, Kafa, and Bench are mainly found in the southwest (Meyer et al., 2023).

The coexistence of multiple languages in Ethiopia creates favourable conditions for language contact phenomena, including code-switching. Code-switching is widely defined as the use of more than one language within a single communicative episode or discourse by bilingual or multilingual speakers (Heller, 2010; Deuchar, 2006, 2012). It typically involves alternating linguistic codes in ways that are systematic rather than random and sensitive to both grammatical and sociolinguistic constraints (Deuchar, 2012; Auer, 2013; Bullock & Toribio, 2009; Gardner-Chloros, 2009; Wei, 2009).

In Ethiopian multilingual communities, CS is a salient feature of everyday interaction and serves important functions in identity construction, stance-taking, and the negotiation of social relations (Leyew 1998). See instances of code-switching in the following examples:

1. *Mastaaweqiiyaa@amh godhee achi kaa'-e factory* ***blue magic@****eng-tu.*
advertisement do-3SL.M.PRV there put-3SL.M.PRV factory **blue magic**-FOC
'It is the blue magic factory that put it there doing advertisement.'

The present study focuses on CS between Afaan Oromoo and English in informal contexts. Afaan Oromoo, a major Cushitic language, plays a crucial role in the cultural and social identity of its speakers (Leyew, 2012). At the same time, English has gained increasing prominence in Ethiopia, especially as a medium of instruction in secondary and higher education and as a language of business and global communication (Bulcha, 1997; Leyew, 2012; Gerencheal, 2019). This expansion has intensified contact between English and Ethiopian languages and has encouraged the incorporation of English lexical items into everyday speech, particularly in urban and semi-urban settings.

Despite widespread bilingualism, there is a notable lack of research on CS between Afaan Oromoo and English, particularly in informal conversation. Existing Ethiopian CS studies tend to centre on Amharic-English or other Amharic-based combinations, and they frequently examine formal domains such as classrooms or media (Ali, 2015; Bejiga, 2021; Sime, 2019; Leyew, 1998). As a result, the everyday CS practices of Afaan Oromoo–English bilinguals, especially at clause level, remain under-described.

Only a few studies have addressed non-Amharic bilingual pairings. Apart from Ali (2015) and Abdisa (2024) who analyses Oromiffa–Harari CS in Dire Dawa, and Afaan Oromoo -Amhric codeswitching, most research neglects the broader linguistic diversity of Ethiopia and the dynamic multilingual practices of communities such as Tigrinya, Somali, and Sidama speakers. This imbalance has produced a narrow account of bilingual experiences and has marginalized major groups such as Afaan Oromoo speakers, despite the numerical and sociopolitical importance of Afaan Oromoo (Meyer et al., 2023).

Ali (2015) remains the only study to directly address CS involving Afaan Oromoo (referred to as Oromiffa in his work). However, his study differs from the present research in several important respects. It investigates CS between Oromiffa and Harari, not English; it focuses on the

Eastern dialect of Afaan Oromoo in the multilingual urban environment of Dire Dawa; and it adopts a mainly sociolinguistic perspective, emphasizing motivations such as identity, solidarity, and communicative efficiency (Ali, 2015). While it offers valuable sociocultural insight, it does not provide an in-depth structural or theory-driven analysis of CS. In contrast, the current study examines the Western Mecha dialect of Afaan Oromoo, widely spoken in western Oromia, and centres on the morphosyntactic and structural properties of CS between Afaan Oromoo and English in informal conversation.

Furthermore, Ethiopian CS studies have rarely applied explicit grammatical frameworks such as the Matrix Language Frame (MLF) model to analyze structural aspects of CS (Myers-Scotton, 1993, 2002, 2004, 2006; Myers-Scotton & Jake, 2017). This methodological gap limits comparability with international CS research and obscures the contribution Ethiopian data could make to general theory. The present study addresses this by adopting the MLF model as its primary framework and by investigating clause-internal switching, morphosyntactic integration, and the distribution of matrix and embedded languages.

Finally, although sociocultural factors, identity construction, and digital communication are clearly relevant to CS in Ethiopia, they require extensive qualitative fieldwork and broader sampling. Given time and resource constraints, this study deliberately restricts its scope to the linguistic dimension of CS. Sociolinguistic variables such as age, gender, education, and proficiency are documented and used descriptively, but they are not subjected to a full variationist analysis.

1.2 Aim of the Study

The aim of this study is to describe and analyze CS between Afaan Oromoo (Maccaa dialect) and English using a relevant structural framework—primarily the Matrix Language Frame model—so that the findings can be compared with CS studies from other bilingual settings (Myers-Scotton, 1993, 2002, 2004, 2006; Myers-Scotton & Jake, 2017). The study focuses on spontaneous, non-institutional interactions and adopts the clause as the main unit of analysis.

Globally, CS has been examined from multiple perspectives, including linguistic, sociocultural, cognitive, and psycholinguistic approaches (Deuchar, 2012; Auer, 2013; Bullock & Toribio, 2009; Wei, 2009; Gardner-Chloros, 2009). In Ethiopia, however, few studies address CS, and the majority focus on Amharic-based pairs or educational contexts (Ali, 2015; Bejiga, 2021;

Sime, 2019; Leyew, 1998). Even fewer employ detailed morphosyntactic analysis or explicit theoretical models, leaving a gap in structurally informed accounts of CS, particularly for Afaan Oromoo.

By concentrating on the Mecha (Maccaa) dialect of Afaan Oromoo and informal convesation, this study aims to fill these gaps. It adopts the Matrix Language Frame model and related concepts such as the 4M model to analyse switching points, integration patterns, and grammatical constraints on CS (Myers-Scotton, 1993, 2002; Myers-Scotton & Jake, 2017). Sociolinguistic variables are recorded but primarily serve to contextualize the linguistic patterns, rather than to support quantitative sociolinguistic generalizations**.**

1.3 Research Questions

The study is guided by the following main research questions:

1. Which language functions as the matrix language in code-switched utterances, and how does this influence sentence structure and word order?
2. How are English lexical items morphosyntactically integrated into Afaan Oromoo speech?
3. To what extent are established theoretical models of code-switching—the Matrix Language Frame (MLF) Model—applicable to code-switching between Afaan Oromoo (Maccaa dialect) and English?

1.4 Hypotheses

Based on previous research and the theoretical assumptions of the MLF and 4-M models, the study advances the following hypotheses:

H1: Afaan Oromoo functions as the matrix language in the vast majority of bilingual clauses, providing word order and system morphemes.

H2: English lexical items are integrated into Afaan Oromoo through systematic, rule-governed morphosyntactic processes that vary according to word class.

H3: The predictions of the MLF model and the 4-M model accurately account for the observed patterns of code-switching between Afaan Oromoo and English.

CHAPTER TWO. THEORETICAL FRAMEWORK AND LITERATURE REVIEW ON CODE-SWITCHING

This chapter presents the theoretical and empirical foundations of the study. It reviews major linguistic approaches to code-switching, with particular emphasis on structurally oriented models that account for grammatical organization in bilingual speech. The chapter justifies the selection of the Matrix Language Frame (MLF) model and the 4-M model as the primary analytical frameworks and situates the present research within the broader literature on code-switching in Ethiopia and beyond. It also reviews existing work on morphosyntactic integration, language attitudes, and transcription practices relevant to the analysis of naturally occurring bilingual data.

2.1 Theoretical Frameworks on Code-Switching

Code-switching has been examined from a wide range of theoretical perspectives, including sociolinguistic, psycholinguistic, and grammatical approaches. Early sociolinguistic research emphasized the social meanings and pragmatic functions of switching between languages (Gumperz, 1982; Auer, 1998, 2013), while later work increasingly focused on the internal grammatical constraints governing bilingual speech (Deuchar, 2012; Deuchar et al 2018; Pfaff, 1979; Poplack, 1980; Muysken, 2000).

From a grammatical perspective, code-switching is understood not as random alternation between languages, but as a rule-governed phenomenon constrained by the syntactic and morphological systems of the languages involved (Pandit, 1990; Myers-Scotton, 1993; MacSwan, 2000). Structural approaches seek to explain how bilingual speakers produce well-formed mixed clauses and how grammatical material from different languages is distributed within those clauses.

Among these approaches, the Matrix Language Frame (MLF) model has been one of the most influential frameworks for analyzing intrasentential code-switching, particularly in spoken data (Myers-Scotton, 1993, 2002, 2006). The present study adopts the MLF model and its extension, the 4-M model, because of their explicit predictions regarding morpheme order, system morpheme selection, and matrix language identification.

2.2 The Matrix Language Frame (MLF) Model

The Matrix Language Frame model proposes that in bilingual clauses, one language—the Matrix Language (ML)—provides the grammatical frame of the clause, while the other language(s)—the Embedded Language(s) (EL)—contribute primarily lexical material (Myers-Scotton, 1993, 2002). According to the model, the ML determines word order, supplies system morphemes, and governs morphosyntactic structure.

A central assumption of the MLF model is that bilingual clauses are asymmetrical: even when elements from more than one language appear, only one language functions as the structural backbone of the clause (Myers-Scotton, 2006). This asymmetry is particularly relevant for analyzing code-switching involving typologically different languages, such as an agglutinative Cushitic language and an inflectionally poorer language like English. The model formalizes its predictions through two core principles: the Morpheme Order Principle and the System Morpheme Principle.

2.2.1 Matrix Language Principle

The Matrix Language Principle states that the matrix language is the language that provides the morphosyntactic frame for bilingual clauses, including constituent order and grammatical relations (Myers-Scotton, 1993). The identification of the matrix language is not based on surface frequency alone, but on structural diagnostics such as the origin of inflectional morphology and functional elements.

In the context of Afaan Oromoo–English code-switching, this principle predicts that if Afaan Oromoo is the matrix language, it should determine clause-level structure, including verb placement, agreement morphology, and case marking, even when English lexical items are inserted.

2.2.2 Morpheme Order Principle (MOP)

The Morpheme Order Principle holds that in bilingual clauses, the surface order of morphemes follows the grammatical patterns of the matrix language (Myers-Scotton, 1993; Myers-Scotton & Jake, 2017). Embedded language elements must conform to the word order constraints of the matrix language in order to be licensed within the clause.

This principle is particularly relevant in contact situations involving languages with different basic word orders. Afaan Oromoo typically follows an SOV order, whereas English follows an SVO order. The MOP predicts that in bilingual clauses where Afaan Oromoo is the matrix language, English lexical items should appear in positions consistent with Afaan Oromoo word order, rather than retaining English syntactic patterns.

2.2.3 System Morpheme Principle (SMP)

The System Morpheme Principle states that system morphemes—such as inflectional affixes, agreement markers, tense, aspect, and case markers—are supplied exclusively by the matrix language (Myers-Scotton, 1993, 2002). Embedded language morphemes are generally excluded from bilingual clauses unless they are lexically required and early system morphemes.

This principle provides a powerful diagnostic for identifying the matrix language and distinguishing code-switching from borrowing. In agglutinative languages like Afaan Oromoo, where grammatical relations are heavily marked morphologically, the SMP predicts that case markers, agreement suffixes, and tense-aspect morphology should consistently come from Afaan Oromoo in bilingual clauses.

2.3 The 4-M Model

The 4-M model is an extension of the MLF framework that classifies morphemes into four types: content morphemes, early system morphemes, late bridge system morphemes, and late outsider system morphemes (Myers-Scotton & Jake, 2000, 2017). This classification is based on differences in how morphemes are accessed during language production.

According to the model, content morphemes (e.g. nouns, main verbs, adjectives) are conceptually activated and can be drawn from either language. Late outsider system morphemes, by contrast, are structurally assigned and must come from the matrix language. This distinction allows for fine-grained predictions about which elements may be switched and which are constrained.

The 4-M model has been successfully applied to a wide range of language pairs, including typologically diverse combinations (Rahimi & Dabaghi, 2013; Khan & Khalid, 2018; Al-Bataineh & Abdelhady, 2019). The present study extends this line of research to Afaan Oromoo, contributing new evidence from an underrepresented language.

2.4 Asymmetry Principle

Closely related to the MLF framework is the Asymmetry Principle, which holds that bilingual speech is characterized by structural dominance rather than equal contribution from the languages involved (Myers-Scotton, 2006; Myers-Scotton & Jake, 2015). Even in contexts of high bilingual proficiency, speakers do not typically combine grammatical systems freely; instead, one language anchors the structure of the clause.

This principle is especially relevant in multilingual Ethiopian contexts, where speakers routinely navigate three or more languages. The present study tests whether asymmetry persists not only between Afaan Oromoo and English, but also in clauses containing Amharic insertions.

2.5 Code-Switching VS Borrowing

It is important to establish clear criteria for differentiating between lexical loans and single-word switches in the code-switching between Afaan Oromoo and English for accurately analyzing language contact. This also offers insights into how external languages influence the structure and vocabulary of Afaan Oromoo over time.

In this study, words in Afaan Oromoo that are listed in authoritative dictionaries will be considered loans, while words not in the dictionary will be categorized as code-switches. This approach helps to clarify the relationship between language mixing and the incorporation of foreign elements in bilingual speech. We have included single-word switches and differentiated them from loans words based on their predictability (Muysken, 2000, p. 71; cf. Deuchar, 2006). This predictability is associated with "listedness," which indicates.

I relied on established dictionaries particularly 'Elelle Bilingual Afaan Oromoo and English dictionary to assess listedness (Hinsen Makuria 2009). If an English-origin term is listed in a recognized Afaan Oromoo dictionary, it is classified as a loan; if absent, it is treated as a switch. We recognized that this approach may not be watertight, as dictionaries may not fully reflect current usage, leading to potential misclassifications. In the context of this study, which aims to identify the matrix language in code-switching instances between Afaan Oromoo and English, loanwords are excluded from the analysis.

This approach is compatible with the Matrix Language Frame model, which predicts varying degrees of morphosyntactic integration of Embedded Language material. As discussed by Stammers and Deuchar, it is important to distinguish between different types of borrowing, including peripheral borrowing, which may exhibit limited integration despite being lexically listed (Deuchar & Stammers, 2012). In line with this distinction, English-origin items in the present dataset that are attested in Afaan Oromoo lexical resources or widely used in otherwise monolingual Oromo discourse are analysed as borrowings, while non-listed items—despite often appearing within an Afaan Oromoo Matrix Language frame—are treated as instances of code-switching.

2.6 Previous Research on Code-Switching in Ethiopia

Despite Ethiopia's multilingual context, research on code-switching beyond Amharic-based pairs is limited, and studies specifically addressing Afaan Oromoo–English code-switching are particularly rare (Balay, 2020; Bejiga, 2021; Sime, 2019; Leyew, 1998; Temesgen & Hailu, 2022). The few existing Ethiopian studies that examine code-switching in educational settings mainly focus on teachers' and students' perceptions and attitudes toward classroom code-switching, rather than on detailed structural analysis.

Temesgen and Hailu (2022) investigate code-switching by EFL teachers in Ethiopia, identifying academic, managerial, and social functions, and showing that code-switching can support comprehension, classroom management, and a positive learning environment. Keleta (2020) analyses Tigrinya–English code-switching in FM radio broadcasts in Mekelle, using the Matrix Language Frame model to demonstrate that Tigrinya functions as the matrix language into

which English nouns, adjectives, verbs, and adverbs are embedded (Keleta, 2020). Emam and Mekonnen (2022) examine Amharic–English code-switching in the media and report frequent insertion of English lexical items into Amharic structures, with Amharic serving as the matrix language (Emam & Mekonnen, 2022).

Although these studies contribute to understanding code-switching in Ethiopia, they differ methodologically and theoretically from the present research and do not provide a systematic grammatical account of Afaan Oromoo–English code-switching. In contrast, the current study focuses explicitly on the structural properties of code-switching between Afaan Oromoo (Mecha dialect) and English, drawing on corpus data and the Matrix Language Frame model as its main theoretical framework.

2.7 Previous Studies on Morphosyntactic Integration

Previous research on morphosyntactic integration in code-switching has largely focused on how embedded-language items are structurally accommodated within the matrix language. The Matrix Language Frame (MLF) model (Myers-Scotton, 1993, 2002) and the related 4-M model (Myers-Scotton & Jake, 2000) offer a key framework for this work, proposing that the Matrix Language (ML) supplies word order, system morphemes, and agreement, while the Embedded Language (EL) contributes mainly lexical items such as nouns and verbs. Morphosyntactic integration is achieved when embedded elements conform to the morphosyntactic rules of the ML.

Studies in typologically diverse contexts broadly support these predictions. In Pashto–English code-switching, Khan and Khalid (2017) show that English verbs and nouns are integrated through Pashto light-verb constructions and inflectional morphology, with Pashto retaining control over tense, agreement, and morpheme order, consistent with the MLF model. Deuchar's (2006) analysis of Welsh–English code-switching similarly demonstrates that Welsh typically provides the morphosyntactic frame, with English content morphemes inserted into Welsh determiner phrases and clause structures; later work confirms that inflection on the finite verb is a reliable indicator of the matrix language (Deuchar, 2006, 2020).

Evidence from other language pairs points in the same direction. Akinrẹmi (2016) reports that in Igbo–English code-switching, English verbs are fully integrated only once they receive Igbo tense and aspect morphology, and all system morphemes and morpheme order follow Igbo, confirming Igbo as the matrix language. Al-Bataineh and Abdelhady (2019), in a study of Cree–English intrasentential code-switching, find that Cree supplies nearly all system morphemes, including case and agreement markers, while English elements function mainly as content morphemes, supporting the MLF Morpheme Order and System Morpheme Principles.

Taken together, these studies demonstrate that morphosyntactic integration in bilingual speech is systematic and strongly constrained by the grammar of the matrix language, especially where the ML is morphologically rich. In such contexts, embedded English items are typically subordinated to the matrix language's word order and inflectional patterns. Building on this comparative evidence, the present study investigates how English lexical items are integrated into the morphosyntactic structure of Afaan Oromoo, an agglutinative, verb-final language with complex verbal morphology and case marking, in spontaneous code-switched conversation. Particular attention is given to word order, system morphemes, and agreement, in order to determine which language functions as the Matrix Language and how structural differences between Afaan Oromoo and English shape patterns of integration (see Chapter 4).

2.8 Practices and Attitudes on Code-switching

Attitudes and practices around code-switching often diverge. Speakers may publicly favour "pure" language use yet rely on mixed speech in everyday communication (Mesthrie et al., 2009; Deumert, 2011). In many bilingual communities, including Cantonese–English and Punjabi–English, code-switching is sometimes criticised but remains a routine interactional resource (Chan, 2022; Chana & Romaine, 1984).

In postcolonial and multilingual settings, evaluations are typically ambivalent: code-switching can index urban sophistication, in-group solidarity, or translocal identity, not just language shift (Stell & Yakpo, 2015; Woolard, 1998). Standard-language ideologies often stigmatize mixed speech as improper even though it is structurally systematic (Bullock & Toribio, 2009).

In Ethiopia, Amharic–English code-switching is widespread in universities, media, and public speeches, with Amharic providing the grammatical frame and English contributing technical and prestigious terms (Leyew, 1998; Getachew Seyoum Woldemariam & Mandefro Fenta Terefe, 2023). Listeners see this as useful for clarity and as a sign of expertise, but some fear it reduces Amharic's status and excludes monolingual audiences (Araya, 2011).

In education, students and teachers generally view strategic code-switching to Amharic as pedagogically helpful, even as some worry about reduced exposure to English (Teklesellassie, 2018; Tamene & Desalegn, 2022). In Oromo–Amharic contact zones, Amharic insertions into Afan Oromo are common yet politically sensitive and linked to debates about Oromo linguistic rights (Abdisa, 2024). Recent work on informal Afaan Oromoo–English interaction shows Oromo as the matrix language, with English and Amharic insertions used for specific domains such as education and technology (Mihiretu & Deuchar, 2025a).

2.9 Transcription and Linguistic Data Analysis

Accurate transcription is central to the analysis of bilingual speech. Following Ochs (1979) and Lapadat and Lindsay (1999), transcription is treated not as a neutral technical process but as a theoretically informed analytical step. The present study adopts an adapted CHAT-based transcription system (MacWhinney, 2000) combined with Leipzig glossing rules (Haspelmath, 2014), allowing for fine-grained morpheme-level annotation.

CHAPTER THREE. METHODOLOGY

3.1 Research Design

This study adopts a qualitative–quantitative mixed-methods design based on naturally occurring conversational data. The aim is to examine the structural organization of code-switched clauses and the morphosyntactic integration of English lexical items into Afaan Oromoo. Naturalistic data are considered the most appropriate source for investigating code-switching because they reflect speakers' unmonitored bilingual practices. The analytical framework is grounded in the Matrix Language Frame (MLF) model and the 4-M model (Myers-Scotton, 1993, 2002; Myers-Scotton & Jake, 2000), which allow for clause-level identification of the matrix language and classification of morpheme types.

3.2 Participants

The participants were 56 bilingual speakers of Afaan Oromoo (Maccaa dialect) and English from the Kellem Wollega Zone of Western Ethiopia. All participants reported Afaan Oromoo as their first language and English as a learned language acquired through formal education. Most participants also reported functional competence in Amharic. The age range of participants was between 21 and 40 years, and both male and female speakers with varying educational backgrounds were included.

3.3 Data Collection

The corpus consists of approximately 21 hours of audio recordings of informal conversations. Recordings were made in natural settings such as homes and social gatherings, and no structured interview questions were used in order to avoid influencing language choice. In addition to audio data, a short sociolinguistic questionnaire was administered to collect background information on participants' age, gender, educational level, language acquisition history, and language use patterns.

3.4 Data Selection and Management

From the full corpus, three recordings were used for full transcription. The selection was guided primarily by analytical feasibility and by the requirement to obtain a sufficient concentration of bilingual material for testing the predictions of the MLF and 4-M models, rather than by an intention to foreground particular interactional contexts.. These recordings yielded 879 analyzable clauses. All data were anonymized and stored in secure digital format. Data management followed a three-stage procedure: (1) audio segmentation, (2) orthographic transcription, and (3) clause-level annotation and coding for quantitative analysis (MacWhinney, 2000; Tagliamonte, 2012).

3.5 Ethical Considerations

Ethical approval was obtained from the relevant institutional body. All participants provided informed consent and were informed of their right to withdraw from the study at any time. Identifying information was removed from transcripts, and all recordings were used solely for research purposes.

CHAPTER FOUR. RESULTS AND DATA ANALYSIS

4.1 Sociolinguistic Profile of Participants

The sociolinguistic background of the participants provides important context for interpreting the code-switching patterns observed in the data. The 56 speakers included in the study are bilingual in Afaan Oromoo (Maccaa dialect) and English, with most also reporting functional competence in Amharic. Afaan Oromoo is the dominant language of daily interaction, while English is primarily associated with formal education and professional.

In terms of age and education, participants ranged from 22 to 40 years and included both secondary school students and university graduates. Most speakers reported acquiring Afaan Oromoo in early childhood and English during formal schooling. Language use patterns indicate that Afaan Oromoo is used predominantly in home and community settings, whereas English is used mainly in academic and institutional contexts. Attitudinal data from the questionnaire suggest generally positive views toward bilingualism, although speakers often perceive code-switching as context-dependent rather than universally appropriate.

4.2 Distribution of Monolingual and Bilingual Clauses

The corpus analyzed consists of 879 clauses extracted from three fully transcribed recordings. These clauses were classified as monolingual or bilingual. The results show that monolingual clauses in Afaan Oromoo constitute the majority of the data, reflecting its role as the primary language of interaction. Bilingual clauses, however, occur frequently and form a substantial proportion of the corpus, demonstrating that code-switching is a regular and systematic feature of informal speech.

Clauses containing English insertions typically involve single lexical items or short phrases embedded within an Afaan Oromoo structural frame. Clauses involving both English and Amharic insertions are less frequent but provide important evidence for multilingual interactional practices in the community. These patterns support previous findings that multilingual speakers draw flexibly on available linguistic resources while maintaining structural coherence.

4.3 Matrix Language Distribution

4.3.1 Monolingual Clauses

In monolingual clauses, Afaan Oromoo overwhelmingly serves as the sole language of production. These clauses follow canonical Afaan Oromoo word order and morphological marking, including subject–object–verb structure and rich inflectional morphology. Monolingual English clauses are rare and typically occur in quoted speech or metalinguistic commentary, reflecting the functional specialization of English in the speech community. See the following example.

(2) Nam-ni amantii isaa sodaat-a.

Man-NOM religion his fear-IMPV

'Man fears his religion.'

(Maccaa-OC01 EYN 544)

(3) He never gives up.

He never give-3SL.M. IMPV up.

'Inni gonkumaa abdii hin kutatu' (equivalent Afan Oromoo translation)

4.3.2 Bilingual Clauses

In bilingual clauses, Afaan Oromoo consistently functions as the matrix language. This is evidenced by the fact that word order and system morphemes such as tense, agreement, and case markers originate from Afaan Oromoo rather than from English or Amharic. These findings are in line with the predictions of the Matrix Language Frame model (Myers-Scotton, 1993, 2002), which posits that the matrix language provides the morphosyntactic frame of bilingual clauses.

English and Amharic appear primarily as embedded languages supplying content morphemes. No instances were found in which English or Amharic supplied the full grammatical frame of a bilingual clause. This distribution supports the Asymmetry Principle, according to which bilingual clauses are structurally dominated by a single language rather than by equal contributions from

multiple languages (Myers-Scotton, 2006; Myers-Scotton & Jake, 2015). The following examples:

(4) **Daily** @eng hin -qaam-u

Daily NEG -chew.1SL.IMP

'I do not chew ‘chat’ **daily**.'

(Maccaa-CO04-GAE 667)

(5) **Attendance**@eng guut-aa-n jir-a.

Attendance fill.CONV.1SL exist-IMPV

'I am filling out attendance.'

(Maccaa-OC016 SIB 96)

(6) *Biri@amh* dhibba lama **transfer@eng** naa godh-i.

Birr hundred two **transfer** **for** me do-IMP.IMPV

'Do **transfer** two hundred birrs to me.'

(Maccaa OC010 YOT 350)

4.4 Morphosyntactic Integration of English Items

4.4.1 Nouns

English nouns are frequently inserted into Afaan Oromoo clauses and are typically marked with Afaan Oromoo case and focus morphemes. They occupy syntactic positions appropriate to Afaan Oromoo grammar and do not retain English plural or possessive marking when integrated. This pattern confirms the prediction of the System Morpheme Principle, according to which system morphemes are supplied by the matrix language (Myers-Scotton, 1993, 2002).

(7) **Photo** @eng’n kan kee duwwaadhaa-?

Photo-NOM that yours alone-COP-Q

‘Is that photo yours alone?’

(Maccaa-OC028 *JIK 6*)

4.4.2 Verbs

English verbs are rarely inserted as bare verb forms. Instead, they are predominantly integrated through light verb constructions, in which an English lexical verb is combined with an Afaan Oromoo auxiliary or verbalizer. This strategy allows the embedded English element to be incorporated into the Afaan Oromoo verbal morphology and tense–aspect system. See the following examples:

(8) Mucatiitu **interview**@eng godha akka yaada isaatti.

Girl.NOM-FOC interview do-3SG.F.IMPRV as idea his--EMPH

'The girl interviews according to his own ideas.'

(Maccaa-OC02-FET-271)

(9) Biri keetiin **pay@eng** gootee nyaatta mitii amma?

Biri your-INS pay do-2SL.CNV.PRV eat-2SG.IMPRV NEG now-Q

"You're eating with the money you paid yourself, right?'

(Maccaa-OC02-SAG-313)

4.4.3 Adjectives and Adverbs

English adjectives and adverbs appear less frequently than nouns and verbs. When they do occur, they are positioned according to Afaan Oromoo word order and may receive Afaan Oromoo agreement or derivational morphology. These items function syntactically as Afaan Oromoo modifiers rather than as independent English structures, further supporting the dominance of Afaan Oromoo in bilingual clause construction. See examples as follows.

(10) Kun ***ikkoo*** **biblically@eng** kan barreefame**.** (Maccaa-OC02-SAG-341)

This then biblically that write-PASS

‘ This is a principle already written biblically’

(11) **Fair**@eng miti.

Fair not.

‘It is not fair.’

(Maccaa-OC- 028 SOG 34)

CHAPTER FIVE. DISCUSSION

5.1 Matrix Language Patterns and Clause Structure

The results demonstrate that Afaan Oromoo consistently functions as the matrix language in bilingual clauses involving English and Amharic insertions. This finding confirms the central claim of the Matrix Language Frame model that bilingual clauses are organized around a single dominant grammatical system (Myers-Scotton, 1993, 2002). In the present data, Afaan Oromoo provides word order, agreement morphology, and case marking, while English and Amharic contribute primarily lexical material.

The dominance of Afaan Oromoo as the matrix language can be attributed to its role as the primary language of everyday interaction in the speech community. Even when speakers draw on English for lexical resources associated with education or modern domains, they continue to rely on Afaan Oromoo grammatical structure. This pattern supports the view that matrix language choice reflects both grammatical competence and sociolinguistic dominance (Myers-Scotton, 2006; Myers-Scotton & Jake, 2015).

The consistent alignment of bilingual clause structure with Afaan Oromoo word order provides strong evidence for the Morpheme Order Principle. Despite the typological differences between Afaan Oromoo (SOV) and English (SVO), English lexical items are positioned in accordance with Afaan Oromoo syntactic patterns rather than preserving English constituent order. This supports previous findings that bilingual speakers do not freely combine grammatical systems but instead integrate embedded elements into a single structural frame (Deuchar 2006; Deuchar et al., 2018).

5.2 Evaluation of the MLF and 4-M Models

The observed patterns of system morpheme distribution strongly support the System Morpheme Principle of the MLF model. Inflectional morphology, including tense, agreement, and case marking, is consistently drawn from Afaan Oromoo rather than from English or Amharic. This confirms that late outsider system morphemes are supplied exclusively by the matrix language, as predicted by the 4-M model (Myers-Scotton & Jake, 2000, 2017).

The classification of morphemes into content morphemes and system morphemes proves analytically useful for explaining why nouns and verbs are more frequently switched than function words. English nouns and lexical verbs are conceptually activated and can be inserted into Afaan Oromoo clauses, whereas grammatical markers such as determiners and tense affixes are constrained by the matrix language. These findings are consistent with studies of typologically diverse language pairs and provide further support for the psycholinguistic motivation underlying the 4-M model (Myers-Scotton, 2002; Myers-Scotton & Jake, 2017).

At the same time, the data highlight the importance of light verb constructions as a strategy for integrating English verbs into Afaan Oromoo morphology. This strategy allows speakers to reconcile differences between the verbal systems of the two languages and illustrates how structural constraints shape bilingual production. Similar integration patterns have been reported in other bilingual contexts involving agglutinative or verb-final languages (Mihiretu and Deuchar 2025a and 2025b), suggesting that this strategy reflects a general typological tendency rather than a language-specific anomaly.

5.3 Comparison with Previous Studies

The findings of this study are broadly consistent with previous research on code-switching in Ethiopian contexts, particularly studies focusing on Amharic–English contact (Leyew, 1998; Keleta, 2020). These studies similarly report that English lexical items are embedded within the morphosyntactic frame of a local Ethiopian language rather than replacing it.

However, the present study extends this line of research by providing a detailed structural analysis of Afaan Oromoo–English code-switching in informal interaction, an area that has received little attention in earlier work (Ali, 2015; Abdisa, 2024). By applying the MLF and 4-M models to Afaan Oromoo, the study demonstrates that these frameworks are applicable beyond Indo-European and Bantu language pairs and can account for code-switching involving Cushitic languages.

The results also converge with international studies showing that code-switching is constrained by grammatical structure and not merely by social or pragmatic factors (Poplack, 1980; Muysken, 2000; Auer, 2013). While pragmatic motivations undoubtedly influence when speakers switch languages, the form that these switches take is shaped by the grammatical architecture of the matrix language.

5.4 Sociolinguistic Implications

The dominance of Afaan Oromoo as the matrix language reflects its strong position in everyday interaction and local identity. English, although socially prestigious and institutionally important, does not replace Afaan Oromoo as the grammatical base of conversation. Instead, it functions as a lexical resource for specific semantic domains, particularly those related to education and modern technology.

This pattern challenges deficit-oriented views of code-switching as evidence of linguistic incompetence. On the contrary, the systematic integration of English and Amharic elements into Afaan Oromoo demonstrates a high level of bilingual control and grammatical knowledge. Code-switching in this community should therefore be understood as a sophisticated communicative strategy rather than as a sign of language erosion.

From a policy perspective, the findings support approaches to bilingual education that recognize the legitimacy of mixed language practices and build on learners' full linguistic repertoires. Acknowledging code-switching as a natural outcome of multilingual competence may contribute to more inclusive and effective language planning in Ethiopia (Leyew, 2012; Meyer et al., 2023).

5.5 Language Ideologies and Code-Switching Practices

Speakers' attitudes toward code-switching reveal a tension between everyday linguistic practice and normative language ideologies. While code-switching is frequently used in informal interaction, it is often evaluated negatively in formal or institutional contexts. This discrepancy reflects broader ideologies that associate linguistic purity with correctness and education, particularly in relation to English.

The structural regularity observed in the data stands in contrast to these prescriptive attitudes. The grammatical patterns identified in this study show that code-switching follows predictable and constrained rules rather than representing random mixing. This mismatch between ideology and practice underscores the importance of linguistic research in challenging stigmatizing views of bilingual speech and promoting a more nuanced understanding of multilingualism ((Woolard, 1998; Auer, 2013).

CHAPTER SIX. CONCLUSIONS

6.1 General Conclusions

This study set out to investigate the structural organization of code-switching between Afaan Oromoo (Maccaa dialect) and English in informal conversational contexts in Western Ethiopia. Using naturally occurring spoken data and adopting the Matrix Language Frame (MLF) model and the 4-M model as analytical frameworks, the study has demonstrated that code-switching in this community is systematic and structurally constrained rather than random.

The findings show that Afaan Oromoo consistently functions as the matrix language in bilingual clauses, supplying word order and system morphemes such as tense, agreement, and case markers. English and Amharic appear primarily as embedded languages contributing content morphemes. These patterns confirm that bilingual clause construction is asymmetrical and anchored in a single grammatical system, in line with the core assumptions of the MLF model (Myers-Scotton, 1993, 2002).

6.2 Theoretical Contributions

The study contributes to the theoretical literature on code-switching by extending the empirical scope of the MLF and 4-M models to an agglutinative Cushitic language in contact with English and Amharic. Previous applications of these models have focused largely on Indo-European and Bantu languages; the present findings demonstrate that their predictions also hold for Afaan Oromoo.

In particular, the consistent alignment of bilingual clause structure with Afaan Oromoo word order supports the Morpheme Order Principle, while the exclusive sourcing of system morphemes from Afaan Oromoo supports the System Morpheme Principle. The data also provide further evidence for the psychological plausibility of the 4-M model's distinction between content morphemes and system morphemes (Myers-Scotton & Jake, 2000, 2017).

6.3 Methodological Contributions

Methodologically, the study illustrates the value of combining naturalistic data collection with fine-grained morphosyntactic analysis. The use of an adapted CHAT-based transcription system together with Leipzig glossing rules enabled systematic clause-level analysis in an under-resourced language.

The three-stage data management procedure—audio recording, transcription, and clause-based coding—offers a replicable framework for future research on bilingual speech in similar multilingual context. The study also demonstrates the feasibility of applying corpus-based quantitative techniques to small-scale spoken datasets.

6.4 Empirical Contributions

Empirically, this study provides the first detailed structural account of Afaan Oromoo–English code-switching in informal interaction. It documents how different word classes are integrated into Afaan Oromoo morphosyntax and shows that English verbs are predominantly incorporated through light verb constructions, while nouns receive Afaan Oromoo case and focus marking.

The inclusion of Amharic insertions further reflects the multilingual reality of Ethiopian speech communities and highlights the importance of moving beyond strictly bilingual models when analyzing language contact phenomena in such settings.

6.5 Limitations of the Study

Despite its contributions, the study has several limitations. The corpus is based on a relatively small number of fully transcribed recordings, which restricts the generalizability of the findings. The focus on informal spoken interaction also means that code-switching in institutional or written contexts remains unexplored.

In addition, the analysis is limited to morphosyntactic structure and does not address phonological or prosodic aspects of bilingual speech. These dimensions may reveal further insights into the interaction between languages in contact.

6.6 Directions for Future Research

Future research could expand the corpus to include a wider range of speakers and communicative settings, including formal and semi-formal contexts. Comparative studies involving other Oromo dialects or other Ethiopian languages would also help to determine whether the patterns observed here are regionally specific or more broadly generalizable.

Further work could integrate discourse-pragmatic and interactional perspectives with structural analysis in order to provide a more comprehensive account of code-switching practices. Longitudinal studies examining changes in bilingual speech over time would also contribute to understanding the dynamics of language contact in Ethiopia.

References

Abdisa, T. (2024). Afan Oromo and Code-switching-Mixing Amharic into Afan Oromo A Case Study of the Relationship between Afan Oromo (Oromo language) and Amharic/Amhara Language in Ethiopia.

Ado, D., Gelagay, A. W., & Johannessen, J. B. (2021). The languages of Ethiopia. Grammatical and Sociolinguistic Aspects of Ethiopian Languages, 48, 1.

Akinrẹmi, I. I. (2016). Phonological adaptation and morphosyntactic integration in Igbo-english insertional codeswitching. Journal of Universal Language, 17(1), 53–79.

Al-Bataineh, H., & Abdelhady, S. (2019). Cree-English intrasentential code-switching: Testing the morphosyntactic constraints of the Matrix Language Frame model. Open Linguistics, 5(1), 706–728.

Ali, N. B. M. (2015). Documentation and Description of Code Switching in Oromiffa and Harari Language. https://api.semanticscholar.org/ CorpusID: 196502523

Ammon, U., Dittmar, N., Mattheier, K. J., & Trudgill, P. (2008). Sociolinguistics/Soziolinguistik. Auer, P. (2005). A postscript: Code-switching and social identity. Journal of Pragmatics, 37(3), 403–410.

Araya, B. (2011). Codeswitching in Tigrinya: The Case of Two FM Radio Programs.

Auer, P. (2013). Code-switching in conversation: Language, interaction, and identity. Routledge. Volume 3 (Vol. 3). Walter de Gruyter.

Balay, B.-E. (2020). Code Switching from Amharic to English in Bilinguals. Accessed: May, 29.

Bejiga, M. (2021). EFL Teachers' Perception and Practices of Code-Switching to Amharic. Language in India, 21(8).

Bulcha, M. (1997). The politics of linguistic homogenization in Ethiopia and the conflict over the status of Afaan Oromoo. *African Affairs*, *96*(384), 325–352.

Bullock, B. E., & Toribio, A. J. E. (2009). The Cambridge handbook of linguistic code-switching. Cambridge University Press.

Chan, B. H.-S. (2022). Translanguaging or code-switching? Reassessing mixing of English in Hong Kong Cantonese. Chinese Language and Discourse, 13(2), 167–196.

Chana, U., & Romaine, S. (1984). Evaluative reactions to Panjabi/English code-switching. Journal of Multilingual and Multicultural Development, 5(6), 447–473. https://doi.org/10.1080/01434632.1984.9994174

Deuchar, M. (2020). Code-switching in linguistics: A position paper. Languages, 5(2), 22.

Deuchar, M. (2006). Welsh-English code-switching and the Matrix Language Frame model. Lingua, 116(11), 1986–2011.

Deuchar, M., & Stammers, J. R. (2012). What is the "nonce borrowing hypothesis" anyway? *Bilingualism: Language and Cognition*, *15*(3), 649–650.

Leyew, Z. (2012). The Ethiopian language policy: A historical and typological overview. *Ethiopian Journal of Languages and Literature*, *12*(2), 1–59.

Meyer, R., Wakjira, B., & Leyew, Z. (2023). *The Oxford Handbook of Ethiopian Languages*. Oxford University Press.

Muysken, P. (2000). *Bilingual Speech: A Typology of Code-Mixing*. Cambridge University Press. https://books.google.hu/books?id=lJI7qrIKmokC

Woldemariam, H., & Lanza, E. (2014). Language contact, agency and power in the linguistic landscape of two regionalcapitals of Ethiopia. *International Journal of the Sociology of Language*, *2014*(228), 79–103.

Woolard, K. A. (1998). *Introduction language ideology as a field of inquiry*. 3–48.