Characterizing multilingual speech patterns in Northern Belize: Profiling code-switches of 18 English-Spanish-Kriol speakers in Orange Walk Town, Belize

Bierings, Emma

Leiden University

Faculty of Humanities

Characterizing multilingual speech patterns in Northern Belize

Profiling code-switches of 18 English-Spanish-Kriol speakers in Orange Walk Town, Belize

Author:

Emma Bierings, MA

S1179187

Thesis supervisor:

Dr. M. C. Parafita Couto

Second reader:

Dr. K. R. Bellamy

A thesis submitted for the

Research master's degree in Linguistics

February, 2025

ABSTRACT

Objectives/research questions: In the present study, we have two main objectives. First, we aim to characterize the proportions of code-switching in the multilingual spontaneous speech of the community in Orange Walk Town, Northern Belize. We analyze overall rates of unilingual versus multilingual speech production and the distribution of the languages used (Spanish, Kriol, and English). To identify the language asymmetry in intraclausal code-switching, we apply the Matrix Language Frame (MLF) Model (Myers-Scotton, 2002), which posits that the Matrix Language (ML) provides the clause's morphosyntactic structure and that system morphemes align with the ML. Previous studies indicate that the speaker's choice of ML in multilingual clauses can depend on various extralinguistic factors (e.g., Parafita Couto, Davies, Carter & Deuchar, 2014). Therefore, this study's second objective is to examine whether a relation exists between the choice in ML and three sociolinguistic factors: (i) self-rated proficiency, (ii) code-switching attitude, and (iii) language use in the participant's social network.

Methodology: We selected and analyzed a subset of nine video recordings from a dataset comprising 50 video recordings of multilingual spontaneous speech from Orange Walk Town. In this subset, eighteen participants (8 females, 10 males, aged 18–25 years, plus one 76-year-old) engaged in paired conversations lasting approximately 15 minutes each (total: 140 minutes). Participants also completed detailed language background questionnaires, providing information on, i.e., proficiency, code-switching attitudes, and language use in their social networks.

Data & analysis: All speech data were transcribed at the clausal level using ELAN transcription software, with further coding and analysis done in Excel.

Results: A total of 4010 clauses were identified, with 3214 (80%) being unilingual, and 632 (16%) multilingual. All participants produced both unilingual Anglophone (English or Kriol) and Spanish clauses within their conversations, indicating habitual engagement in multilingual speech and code-switching within and between clauses or full sentences. Among the unilingual clauses, 1921 (60%) were Anglophone (English or Kriol), slightly surpassing the 1291 (40%) Spanish unilingual clauses. Although no considerable language preference was observed in unilingual clauses, there was a marked preference

for Spanish as the ML in multilingual clauses; with 539 (87%) being Spanish ML versus 83 (13%) Anglophone ML. Proficiency levels related to language preference for unilingual outcomes but not for multilingual outcomes. The preference for Spanish ML in multilingual clauses may be related to the prestigious status of Kriol. Further exploration of the speaker's social network data is needed to relate this factor to speech production outcomes.

Originality: This study presents new insights by analyzing recently collected spontaneous speech data from a young generation in a relatively unexplored multilingual community. The findings shed light on how proportions of different languages are distributed in multilingual speech, shaped by community norms or individual sociolinguistic factors.

Implications: The study reaffirms that code-switching is a common practice in Northern Belize. Despite individual differences in the proportions of unilingual versus multilingual speech, the choice of ML in multilingual speech reflects a community-wide norm. This ML choice, which inherently involves an asymmetry in language dominance within a clause, can be influenced by sociolinguistic factors, such as switching toward a language with higher prestige, as found by Blokzijl, Deuchar and Parafita Couto (2017). Future research should delve deeper into this Northern Belizean community on a larger scale, considering variables such as age and individual language attitudes toward Kriol, Spanish, and English. Additionally, examining code-switching patterns within conflict sites, such as determiner-noun-adjective constructions, could enhance our understanding of how code-switching operates and whether community standards influence these patterns.

De stilte voor de storm...

ACKNOWLEDGEMENTS

First, I would like to thank my thesis supervisor, M.C. (M. Carmen) Parafita Couto (Leiden University Centre of Linguistics, LUCL) at Leiden University. M. Carmen has known me for quite some years now. She was the second reader of my bachelor's thesis in Latin American Studies and my thesis supervisor for the master's of the same program. After my graduation in Latin American Studies in 2020, I followed another path than in academics. In March 2024, I walked into her office, and she immediately accepted to be my supervisor, once again. For this thesis, M. Carmen invited me to collaborate on the Crossing Language Borders project that she is affiliated with. She introduced me to the rest of the project team and involved me in an enjoyable and inspiring research environment. Throughout the thesis research and writing process, M. Carmen was always available to tackle my doubts and answer my questions, even outside office hours (once again, I apologize, M. Carmen!). With her incredible knowledge and expertise in the field, she helped me move forward. Her ambitions inspired me and always encouraged me to stay curious and ask the essential, critical, and crucial questions in (code-switching) research. She let me gain confidence in my work and helped me grow academically and personally. I am immensely grateful for that and cannot possibly thank her enough.

I would also like to thank the whole research team of the Crossing Language Borders project for the inspiring support, and two members in particular. First, I thank Renzo Ego Aguirre Santa Cruz, who collected the data in Northern Belize and let me use it for my study. I also thank him for answering my many questions about the data collection process and providing me with all the necessary background information and documents. Without this, I would not have been able to build the corpus for this study and do the research I wanted to carry out. My gratitude also goes to Dr. Floor van den Berg, who provided me with the complex yet beautifully displayed network data. Thank you for promptly sharing this so I could use it to complement my outcomes.

I would also like to acknowledge the great team and the immense help in the transcription and coding process. First, I would like to thank Renzo, once again, and Kiki Zanolie for helping me transcribe several video recordings that I included in the analyses. These

videos were sometimes hard to understand, especially when there were languages involved that we were not that familiar with. I am aware of this challenge, and you both managed to work through this. I am also grateful for Simon Claassen, who meticulously checked several transcribed files for inconsistencies, partly cleaned out the unrecognizable speech, and reviewed the organization of the clauses in these files. I would like to thank Luisa Rini as well, who helped me organize the transcripts imported into Excel at a later stage and helped me code part of the files. I am well aware that transcribing and coding takes many hours, and building a corpus is not a one-woman's job, so thank you all -immensely- for helping me out.

I am also grateful for an expert in the field of Northern Belizean code-switching; Dr. Osmer Balam, who was occasionally confronted with my doubts or findings in this study. I want to thank him for the interest, openness, and kindness he was responding to those, and for the valuable feedback he gave me on these matters.

Finally, I must express my profound gratitude to my friends and family for their continuous encouragement throughout this thesis process. I am especially gratefully indebted to my dearest partner, Rob van der Heijden, with his unfailing support. He supported me wholeheartedly and patiently throughout the thesis process by listening to my insecurities and brainstorming during the research. He always encouraged me to take one step at a time and stay curious. He reminded me, more than once, that enjoying the research and writing process is the most important part of it all. I cannot think of words that cover the gratefulness I cherish for him.

Each of your advice, endless encouragement, and positivism made this thesis feel like an achievement to be proud of. This accomplishment would not have been possible without any of you. Thank you.

Emma Bierings

ABBREVIATIONS

All linguistic abbreviations in this thesis are according to the Leipzig Glossing Rules (Comrie, Haspelmath & Bickel, 2015). Our examples show Spanish in **bold**, Belizean Kriol in *italics*, and English in regular.

1	first person
2	second person
3	third person
AD	adjective
AR	article
AUX	auxiliary
DAT	dative
DEF	definite
DEM	demonstrative
DET	determiner
DIM	diminutive
DIST	distal
F	feminine
INDF	indefinite
M	masculine
PL	plural
POSS	possessive
PFR	perfect
PRS	present
PROG	progressive
PST	past
SG	singular
SBJV	subjunctive

Table of Contents

1. Introduction

Code-switching (hereafter CS) is a common practice among bilinguals and multilinguals. It is the alternation between two or more languages in a single conversation, in which the speech is governed by grammatical constraints of at least one of the languages involved (e.g., Bullock & Toribio, 2009). The point in the conversation where the languages change, so-called switch points, are selected by the speaker automatically, although not arbitrarily (e.g., The Equivalence Constraint, Poplack (1980)). These switch points can appear in different parts of speech. Example 1 (**Spanish**/English) illustrates a sentence with multiple switch points where code-switching occurs.

1. '**El**   package   **fue**   delivered   **esta mañana'**

   The   package   be.3SG.PRET   delivered.PSTPART   this morning

   "The pack was delivered this morning."

   (Balam, Stadthagen-González, Rodriguez-González & Parafita Couto, 2022:417)

The practice of CS can occur in a small group of individual speakers (e.g., in families of immigrants in the U.S.A. (Baran, 2017)), or it can be the day-to-day practice within an entire bi- or multilingual or bimodal speech community (e.g., in the multilingual speech community in Northern Belize (Balam, 2014)).

  The present study examines the overall language proportions and distribution of CS in spontaneous speech within the multilingual community of Orange Walk Town, Northern Belize. Additionally, it explores sociolinguistic factors that may be related to the observed multilingual speech patterns. Owing to a complex history involving Belize and neighboring countries such as Mexico (to the north) and Guatemala (to the west and south), Belize is home to a diverse range of ethnic groups and languages. Even though the official language of Belize is English, Kriol or Spanish are the *lingua franca*, the predominant spoken language, depending on the region. In Northern Belize, the *lingua franca* is Spanish, with Kriol and English as the other most common languages (Balam, 2016a). Notably, in Northern Belize, CS is a routine and socially embedded practice, with

generally positive community attitudes toward its use (Balam, 2013). Language attitudes can significantly influence whether multilinguals engage in CS. Instances where one of the languages involved in CS is perceived negatively, can lead to an indirect decrease in CS. This is seen in the case of Llanito in Gibraltar, for instance, where negative attitudes toward Spanish reduce the use of English-Spanish CS (Macdonald, 2024). Other factors, such as language proficiency levels or language use in the social network of the speakers can also influence CS (e.g., Parafita Couto, Davies, Carter & Deuchar, 2014). Exploring the proportions of unilingual and multilingual speech patterns of this Belizean community, alongside CS attitudes, proficiency levels and possible other factors influencing multilingual outcomes, provides valuable insights into theoretical approaches that seek to explain the grammatical underpinnings of CS practices.

Over the past decades, numerous theoretical approaches emerged in CS research (e.g., lexicalist and usage-based approaches), which intend to explain and predict outcomes of multilingual speech patterns (Parafita Couto, Bellamy & Ameka, 2023). In this study, we mainly focus on one of the widely applied theoretical approaches in the field of CS research: the Matrix Language Frame (MLF) model by Myers-Scotton (2002). This model builds on Joshi's (1985) observation that bilingual speech exhibits an asymmetry between the two languages, with one functioning as the dominant matrix language (ML) and the other as the embedded language (EL) (cf. Parafita Couto et al., 2023). Myers-Scotton (2002) extends this observation into a framework explaining the structural grammatical constraints underlying this asymmetry.

The model states that the Matrix Language (ML) will provide the morphosyntactic frame of the clause, to be observed by system morphemes (e.g., determiners, pronouns, and inflectional morphemes). At the same time, the Embedded Language (EL) needs to follow this frame strictly and can, therefore, only contain non-conflicting content morphemes within the same clause (e.g., inserted nouns, verbs, and adjectives). Determining the ML-EL hierarchy in a clause, is realized by two principles: The Morpheme Order Principle, and the System Morpheme Principle. The Morpheme Order Principle states that in ML-EL constituents consisting of singly occurring EL lexemes and any number of ML morphemes, surface morpheme order will be that of the ML (Myers-Scotton, 2002:59). The System Morpheme Principle states that in ML-EL constituents, all

system morphemes which have grammatical relations external to their head constituent, will come from the ML (Myers-Scotton, 2002:59). As an exception, embedded language islands (Embedded Islands) can occur within the ML and can violate the rules of these Principles. We will explain the principles and the embedded islands by using some examples. For instance, Examples 2 and 3 illustrate the Principles of Myers-Scotton's MLF model in a code-switched English-Spanish clause.

2. 'She wears a beautiful **vestido**.'   3. She wears a **vestido**   **bonit-o**.

            dress.M.SG   dress.M.SG   beautiful-M.SG

"She wears a beautiful dress"   "She wears a beautiful dress"

In both cases, English serves as the Matrix Language (ML), as indicated by elements such as verb inflection, pronouns, and determiners being in English. Conversely, Spanish functions as the Embedded Language (EL), with the inserted noun in both examples, and the adjective in Example 3, being in Spanish. In English, the noun-adjective word order is prenominal 'beautiful dress', and in Spanish postnominal 'dress beautiful'. However, it is essential to note that in Example 3 the clause follows the Spanish noun-adjective word order, '**vestido bonito**'. According to the MLF model, this is only possible if the abovementioned Principles are not violated, with one exception. In this case, the isolated '**vestido bonito**' is an Embedded Language Island, violating the Principles by not following the word order of the ML (English). Embedded Islands within the clause can follow the grammatical rules of the EL. In this case, the EL word order and inflected grammatical gender- and number agreement are according to Spanish grammar.

Besides the grammatical explanations of CS and exploring the answers to the questions of *when* and *how* CS occurs, we also ask ourselves *why* bilinguals and multilinguals code-switch. According to Gardner-Chloros (2009:98), the answer lies in the study of social factors regarding CS: "*CS is in fact a construct that is derived from the behavior of bilinguals*", and in continuation, she argues "*Although (...) CS is now studied from a number of different perspectives and with different methodologies, the primary source of data remains in the sociolinguistic arena*". Gardner-Chloros (2009:98-99) also

describes a wide range of factors that can determine whether or not code-switches occur in discourse. She divides these factors into three types: (i) Factors independent of the speakers and circumstances, such as prestige, (ii) factors directly related to the speakers, such as their social identity, attitudes, and ideologies, and (iii) factors within the conversations where CS takes place, such as CS ability of the conversational partner. Many of these factors overlap and interact with one another. Besides factors such as age, gender, proficiency, language attitudes, age of acquisition, the sense of ethnic or religious belonging, and social status, and also the lesser explored factors, such as personality and interpersonal relationships, influence the practice of CS (Dewaele & Li, 2014). This great variety of factors can be investigated on a macro-level, which pays attention to factors as a consequence of institutional and social boundaries within the community. It can also be explored on a micro-level, which places the focus on the conversations themselves, the speakers within those, and their social motivations behind their speech behavior (Nguyen, 2015:17). Even though CS is a practice done by the individual speaker, the interlocutors in the conversation as well as the entire speech community influence the outcomes between the bilingual or multilingual speakers.

Due to various factors influencing the practice of CS, several social motivations have been explained on a macro-level as to *why* individuals code-switch (and *why not*). For instance, one of the aspects that Blokzijl, Deuchar, and Parafita Couto (2017) investigated was the directionality of code-switches between Nicaraguan Creole English (NCE) and Spanish in Nicaragua. The authors examined determiner-noun switches in relation to the Matrix Language of the multilingual clause. Assuming an asymmetry between these languages, the MLF model posits that either Spanish or NCE serves as the ML – the dominant language - in a clause. Consequently, one of the two languages may be preferred over the other as the ML. Analyzing the choice of ML and, therefore, examining the preferred direction of the language switch (from NCE to Spanish or vice versa) provides insights into whether there are systematically preferred community norms. Remarkably, Blokzijl et al. (2017) found that all determiner-noun switches were from NCE into Spanish, and not vice versa. The authors suggest that the switch occurred in the direction of the language with the higher prestige. Given that Spanish is the official

language in Nicaragua, it is considered to have higher prestige than NCE (Blokzijl et al., 2017:12).

Socially embedded reasons to code-switch at the community level also include fostering a sense of inclusion. For instance, in Ghana, an Akan radio station uses CS between English and Akan on a talk show to disseminate information to listeners (Brobbey, 2015). Typically, only one language is broadcast on this radio station. However, when bilingual hosts engage in conversations on particular radio talk shows (e.g., on sports or politics), they are allowed to speak in their habitual mixed speech. This way, CS becomes a tool for a wider spread of information. Despite the possibility that not all the information is discussed in both languages (and not all listeners may understand both languages), the listeners can still grasp the core of the story (Brobbey, 2015). Similarly, in Northern Belize, bilinguals or multilinguals experience a sense of inclusion or belonging when they code-switch, as it is a habitual form (and a social norm) of speech within the community (Balam, 2014; Fuller Medina, 2021).

The differing attitudes toward CS and the various reasons for individuals to engage in this linguistic practice, contribute to the emergence of community-specific norms regarding the extent, manner, and sociolinguistic implications of CS. This study seeks to examine and delineate CS practices of multilingual speakers in Orange Walk Town, Northern Belize, through the analysis of their naturalistic speech and the sociolinguistic attributes of the community. Previous research by Balam (2014, 2016a) has identified CS involving two or more languages as a routine feature of daily communication within this speech community. For this reason, we pay particular attention to sociolinguistic factors. More specifically, we concentrate on factors such as CS attitudes, language proficiency levels, and language use in the speaker's social network. From a grammatical perspective, this study examines the overall distribution of unilingual and multilingual clauses, followed by a focused analysis of the language use in intraclausal code-switches. This analysis provides a detailed account of the languages involved and the choice of the ML.

This investigation enhances our understanding of how multilingual speakers engage within a community where CS is a normalized and integral part of daily communication, revealing the specific patterns that emerge in such contexts. Studies

addressing CS in multilingual settings with naturalistic data and encompassing entire communities where CS is a habitual practice remain scarce. Consequently, the findings of this study offer valuable insights into the mechanisms of CS in multilingual environments and contribute to refining theoretical frameworks, such as the MLF model, that aim to explain CS patterns.

We believe perhaps that's enough of Chapter 1.[1] In the following chapters, we present the details of our study. In Chapter 2, we outline several theoretical approaches to the study of CS, with particular emphasis on Myers-Scotton's MLF model (2002). This chapter explores the applications of the MLF model and reviews key findings from previous literature. Chapter 3 provides a historical overview of multilingualism in Belize, with a focus on the languages relevant to our research- Kriol, English, and Spanish - along with an analysis of the unique characteristics of each language. Continuously, it examines the social status and attitudes associated with these languages within the community. Subsequent chapters focus on the empirical data from this study. Chapter 5 details the research methodology and the process of constructing a corpus of multilingual spontaneous speech. It discusses the data collection in Northern Belize and provides insights into the participants, including their metalinguistic profiles. The transcription process and data coding are described in preparation for analysis, which is conducted in Chapters 6 and 7. Chapter 8 summarizes the study's findings, reflecting on the theoretical implications of the MLF model in light of our results. In the final chapter (Discussion), we consider the broader implications of the study, connecting extralinguistic and sociolinguistic factors to the findings. We also explore how this research contributes to the field of CS studies and propose directions for future research.

---

[1] Parafita Couto (2005:33).

2. Grammatical approaches to CS

This chapter seeks to provide grammatical explanations for the occurrence and mechanisms of CS. We begin by describing the various types, strategies, and categories of CS. Subsequently, we explore the theoretical approach of the Matrix Language Frame MLF model by Myers-Scotton (2002) and discuss its widespread application by previous scholars. We will further set out a grammatical conflict site (where grammars intersect (Poplack, 1980)) in multilingual noun constructions involving determiners and adjectives. In continuation, we briefly review relevant studies in the area of Northern Belize, their findings, and the grammatical outcomes observed. Finally, we will review research methods in CS research, with a focus on building a multilingual corpus of naturalistic speech, the research method of our study.

2.1 Types of CS

One of the foundational studies in CS research is Poplack's (1980) comprehensive examination of CS structures and grammatical constraints in naturalistic speech among Spanish-English bilinguals in Puerto Rico. Poplack (1980) distinguishes between two types of CS in her grammatical approach: extrasentential and intrasentential CS. Extrasentential CS occurs when the speaker switches from one language to another between two sentences, as illustrated in Example 4, with a code-switch from English to **Spanish**.

4. 'I think we're gonna go back to Paraíso. **Allá   visitar   a   la   cas-ita.'**

            There   visit.INF   to   DEF.F.SG   house-DIM.F.SG

  "I think we're gonna go back to Paraíso. There, to visit the house."

        (participant_2_adrtun in adrtunelamog171223ow)

Included in full-sentence switches, Poplack (1980:596) added a category with *fillers* (e.g., **este** 'uhm', I mean), *interjections* (e.g., **¡ay, Dios mío!** 'oh, my God!', shit!), *idiomatic expressions* (e.g. **y toda esa mierda** 'and all that shit', no way), *quotations* (e.g., put down **menos** 'less'), and tags (e.g., you know, **¿entiendes?** 'you understand?'). She argues that

these segments can occur at any point in the sentence and do not adhere to sentence-internal syntactic constraints. Therefore, this category falls within extrasentential CS and is considered 'less intimate' than other forms. This contrasts with Poplack's (1980) second type of CS, intrasentential switches, which occur within sentences and are deemed a more 'intimate' form of switching between languages. An additional, somewhat outlying category within intrasentential CS includes the insertion of single nouns into otherwise unilingual sentences. While Poplack (1980) hypothesized that these single nouns are commonly heavily loaded with ethnic content, her results only partly supported this hypothesis. She also noted that previous studies found noun insertions (of the non-dominant language) to be the most common type in intrasentential switching. Her findings aligned with these observations and were later corroborated by researchers such as Muysken (2000), Matras (2009) and also supported by Fuller Medina (2021) on English-Kriol-Spanish CS in Northern and Western Belize. On the topic of the term 'intrasentential CS', Deuchar (2012) further clarifies that intrasentential CS can refer to both clause switches within a sentence and switches within a clause, while interclausal and intraclausal CS specifically indicate that the switch point lies between two clauses of the same sentence (Example 5), or within the clause (Example 6), respectively.

5. 'When is the time   **que   vas   a   ir   a   Chetomal?'**

                            that   go.2SG   to   go.INF   to   Chetomal

  "When is the time that you will go to Chetumal?"

                            (participant_1_elamog in adrtunelamog171223ow)

6. '**Bueno,   tu**   grandmother   **est-á**   happy'

    Well,   2SG.POSS   grandmother   be-2SG.PRS   happy

  "Well, your grandmother is happy"

                            (participant_1_elamog in adrtunelamog171223ow)

The switch point(s) of the most commonly studied type of CS, intraclausal CS, as demonstrated in example 6, provide insights into how bilinguals or multilinguals

linguistically behave when two or more grammars have underlying conflicting grammatical and structural constraints. For this study, we use Poplack's (1980) term *extrasentential CS* for CS in between complete sentences, as in Example 4. However, we adopt Deuchar's (2012) distinction between *interclausal* and *intraclausal CS* (as in Examples 5 and 6), to avoid confusion regarding the specific point in the sentence where the code-switch occurs.

When categorizing types of CS, we not only consider the place in discourse where CS occurs, as described above, but also distinguish the operating strategies of switches. These categories include discourse markers and single-noun insertions mentioned previously. For instance, Poplack (1980) noted that a bilingual clause can contain one embedded word or speech segment (such as compounds or compound expressions) in an otherwise unilingual clause. Muysken (2000) divides these code-switch strategies into three categories: *insertion*, *alternation*, and *congruent lexicalization*. Muysken (2013) adds a fourth category: *backflagging*. Below, we briefly describe each strategy, noting that these are based on bilingual rather than multilingual practices of CS. We base the explanation of each category on Muysken's (2013) paper, which is an updated version of Muysken's (2000) book.

Insertion: the insertion of language B (mostly single or compound lexical elements) into a dominant language A. In example 7, we illustrate this type of switch, in which we consider 'church field' as a compound noun and 'December' as a single word insertion in an otherwise Spanish-dominant clause. The participants in this conversation talk about work-outs before the festive days in December:

7. **'Vueltas**   **en la**   **c- en** church field **para**   **tener**   **un**
go.back.2SG   in   DEF.F.SG   c- in   for   have.INF   INDF.M.SG

mejor   **cuerpo**   **para**   **est-**   **este**   December?'
better   body   for   est-   DEM.M
"Do you go back to the church field to have a better body for December?"

                   (participant_1_jacbla in jacblakiamai211123ow)

Alternation: The switch is more discourse-oriented and can occur independently of the grammars involved. In example 8, 'then' clearly is discourse-oriented, and 'back' can occur independently of both grammars.

8. 'Then **regres-a**   back **aquí, a Orange Walk'**
         return-3SG   here   to   Orange Walk

  "Then he/she returns back here, to Orange Walk"

                                   (participant_2_adrtun in adrtunelamog171223ow.eaf)

Congruent lexicalization occurs when languages A and B share structural properties with two typologically closely related languages. Or when two languages are not typologically related but when these grammars structurally agree. In example 9, the verbal word order of English and Spanish agree in '**vamos a estar** united', 'we are going to be united', with 'united' as the final element in the verb phrase. This means CS can occur as congruent lexicalization based on this grammatical rule.[2]

9. '(...) **para** Christmas **va-mos   a   estar**   united, **claramente'**
        for   go-1PL.PRS to   be.INF   clearly

     "for Christmas, we are going to be united, clearly"

                                   (participant_1_elamog in adrtunelamog171223ow)

Backflagging: Clause peripheral (discourse) markers concern single items and are simple and frequent. These markers usually have a clear ethnic connotation. Backflagging is common among heritage speakers, who use their learned L2 as a dominant language and insert backflagging markers from their heritage L1. Example 10 shows instances of *backflagging* in a **Spanish** dominant clause with *Kriol* and English discourse markers. Note, for this example, we do not argue about whether Kriol or English are heritage

---

[2] We are aware that the grammar does not wholly agree in this case as Spanish 'united' (in this case *unidos/unidas*) also usually marks number- and gender agreement of the subject (*we*). However, Muysken (2013:713) describes that in CS with congruent lexicalization, grammars can be wholly or *partly* shared.

languages; we emphasize the possible ethnic connotation with discourse markers, peripheral to the Spanish clause. Our next chapter elaborates on this possible connotation.

10. '*Bwai*, **ese** **día,** **quem-é** **mi** **mano**, like...'
    man DEM.DIST day burn-1SG.PST POSS.1SG hand like
    "Man, that day, I burned my hand, like..."

$$\text{(participant\_2\_shecar in emmarcshecar181223np)}$$

Our study will examine intraclausal CS in naturalistic speech, which will include insertions, alternation, congruent lexicalization, and backflagging. Linguistic units such as *fillers*, *interjections*, *idiomatic expressions*, *quotations*, and *tags,* as described by Poplack (1980) are henceforth treated as discourse markers[3] peripheral to the clause.

2.2 Reviewing and applying the MLF model

CS has been extensively studied through various theoretical frameworks, each offering unique insights into the mechanisms and motivations behind this phenomenon. For a comprehensive overview of the mayor theoretical approaches, along with critical remarks on the applications of these perspectives, see e.g., Parafita Couto et. al. (2023).

Our study adopts the widely applied Matrix Language Frame (MLF) theoretical model, proposing that language use in multilingual contexts, resulting in CS, is asymmetrical. We draw upon Myers-Scotton's (1997, 2002) assumption that when intraclausal CS occurs, only one language dominates over the other(s). Myers-Scotton (1997, 2002) argues that the dominant language, the Matrix Language (ML), provides the morphosyntactic frame of the clause, to be observed by system morphemes (e.g., determiners, pronouns, and inflectional morphemes). Meanwhile, the other language(s), the Embedded Language(s) (EL(s)), need(s) to strictly follow this frame and can only contain non-conflicting content morphemes within the same clause (e.g., inserted

---

[3] Fillers are often used as discourse markers and are language specific, see Crible, Degand & Gilquin (2017).

nouns, verbs, and adjectives). As explained in Example 3 in the introductory chapter, Embedded Islands are insertions within the morphosyntactic frame of the clause that follow the grammatical rules of the Embedded Language of the Island.

The application of the MLF model on naturally produced multilingual speech can go as follows. If we analyze Example 11 according to the MLF model, we observe that the ML is Spanish; verb inflection (**-o**) and determiners (**un**) are in Spanish. Spanish is also a pro-drop language, meaning the first-person singular pronoun is embedded in the verbal inflection. These are three indicators that the ML is Spanish. Consequently, according to the MLF model, all system morphemes are in Spanish and follow the morphosyntactic frame of the ML. Simultaneously, we can observe that the English 'following year,' 'scholarship', and 'U.S.' are embedded within this frame.

11. '**Sí,** **tal vez** **en** f- uhm following year, **esper-o** **conseguir**
      yes perhaps in f- uhm following year hope-1SG.PRS get.INF

   un **un** scholarship, **para** **allá** **en** U.S.'
   DEF.M.SG DEF.M.SG scholarship for there in U.S.

"Yes, perhaps in the following year, I hope to get a scholarship, for there in the U.S."

<div align="right">(participant_1_emaort in carchaemaort171223ow)</div>

The model accounts for this result by positing that the EL contains non-conflicting content morphemes, such as adjectives and nouns. As the EL must adhere to the grammatical rules of the ML, we observe that (i) 'scholarship' and 'U.S.' are inserted nouns that do not violate any syntactic Spanish grammatical rules (those of the ML), and (ii) 'following year' presents a different word order than the usual Spanish order; adjectives are typically postnominal. However, if we consider 'following year' as an Embedded Island, it should follow the grammatical rules of the EL. In this case, the EL is English, and the adjectives are in a prenominal position. Therefore, we conclude that in example 11, the clause aligns with the predictions of the MLF model, and the grammars

of languages in the clause are asymmetrically divided, with Spanish being the dominant language over English.

Over the past decades, numerous studies have carried out CS research based on this MLF model, accounting for and evaluating the results of the studies, and making predictions on CS outcomes. The MLF model has been profoundly attested in CS studies involving different language pairs, the same language pair across different communities, with children, adults, and heritage speakers' speech, and with both naturalistic and experimental data[4]. Although not all outcomes can be explained by the MLF model, it remains a well-established theoretical approach to explaining and predicting CS practices.

2.2.1 Previous studies attesting the MLF model

The MLF model and its assumed language asymmetry are particularly intriguing at grammatical conflict sites. For instance, in determiner-noun-adjective constructions, Spanish exhibits gender and number agreement on both the determiner and adjective, with adjectives typically appearing postnominally relative to the noun (see Example 12). In contrast, English determiners and adjectives do not reflect gender or number agreement with the noun, and adjectives are prenominal (see Example 13).

12. '**los   vestido-s   bonit-o-s'**   13. 'the   beautiful-ø-ø   dress-es'

   DEF.M.PL   dress.M-PL   beautiful-M-PL   DEF.ø.ø   beautiful-ø-ø   dress.ø-PL

   "the beautiful dresses"   "the beautiful dresses"

---

[4] In between countless others, but for each variety of study, see e.g., Deuchar, 2006; Huddlestone & Nel, 2012; Rahimi, 2013; Ihemere, 2016; Eppler, Luescher & Deuchar, 2017; Fairchild & Van Hell, 2017; Parafita Couto & Stadhagen-González, 2017; Balam & Parafita Couto, 2019; Bierings et al., 2019; Forker, 2019; Królikowska, Bierings, Beatty Martinez, Navarro Torres, Dussias & Parafita Couto, 2019; Al-Bataineh & Abdelhady, 2020; Kniaź & Zawrotna, 2021; Vanhaverbeke & Enghels, 2021; Khan, Khalid & Saleem, 2023; van Osch, Parafita Couto, Boers & Sterken, 2023; Torres Cacoullos & Vélez Avilés, 2023.

Among many others examining this conflict site, Bierings et al. (2019) analyzed *Kaqchikel* (Mayan)-**Spanish** nominal constructions involving determiners and adjectives. *Kaqchikel* follows the same grammatical rules as in English: no gender or number agreement is reflected on the determiner or adjective, and adjectives are usually prenominal. The authors attested this with elicited nominal constructions using the Director-Matcher Task (see Gullberg et al., 2009). They defined the ML based on the occurrence of system morphemes in these constructions (e.g., by possessive prefixes, see Example 14). They concluded that the word order of the ML was omitted due to a task effect (or a coping strategy), rather than finding outcomes in grammatical solutions for this conflict site.

14. '*Jun*   **círculo**   *ru-b'onil*   *säq'*

     INDF   circle   3SG.POSS-color   white

    "a circle, colored white"

<div align="right">(DMT-19, P1 in Bierings et al., 2019:11)</div>

In addition to elicited data, corpus-based studies on naturalistic speech data have utilized the MLF model to establish CS patterns within and across different communities. For instance, Blokzijl et al. (2017), examined the language of the determiner in Nicaraguan Creole English (NCE)-Spanish mixed bilingual nominal constructions and found a preference in choice for the language of the determiner, thus the ML. They attributed this preference to a sociolinguistic factor, specifically the language of prestige, which in this case was Spanish. Carter et al. (2011) investigated the ML asymmetry in three different bilingual communities: Welsh-English (Wales), Welsh-Spanish (Patagonia), and Spanish-English (Miami), each with different outcomes. They explained the uniformity of ML usage (Welsh preference) in both unilingual and bilingual clauses in Wales and Patagonia by community norms. In Miami, the choice of the ML was more diverse. While the unilingual ML was almost equally divided between Spanish and English, Spanish was preferred as the ML in about two-thirds of the bilingual clauses. The authors explain these differences based on varying proficiency levels, social networks, and ethnic identities of the speakers.

Other studies have also reported different preferences within and across communities. For example, Królikowska et al. (2019) examined Spanish gender assignment of the determiner in elicited Spanish-English Determiner Phases (DPs) through a Map Task (see Gullberg et al., 2009) across four different communities. They specifically analyzed gender assignment strategies in DPs of individual speakers and related the outcomes to speakers of the same community. Remarkably, they found similar patterns (preferences) within each community and different patterns (preferences) across the four communities. For instance, one community (Puerto Rico) preferred using the Spanish default gender (masculine) over the non-default (feminine) assignment of the determiner. In another community (El Paso), they preferred assigning the gender of the determiner based on the Spanish translation equivalent of the (gender-related) noun.

In conclusion, when multiple grammatical outcomes are possible at a conflict site, entire communities may develop a norm or preference for one outcome over the other, indicating a pattern of uniformity (such as the choice of the ML in bilingual speech or strategies in grammatical gender assignment). However, these patterns may vary on an individual level or across communities. Explanations for these preference variations may be rooted in social motivations.

2.2.2 Applying the MLF model in Northern Belize

In the present study, we seek to determine whether such a community norm exists in Orange Walk Town in Northern Belize and whether there is a ML preference for one language over the other in our dataset. That being the case, we further explore sociolinguistic factors that may explain these community patterns.

Most research on the languages in Belize until the late 1990s included studies on language attitudes, ideologies, and related ethnic identities. This is not surprising, as Belize gained independence from Britain in 1981 and research focus lied on, i.e., identifying being 'Belizean' by speaking Kriol (Decker (2013), further explained in our next chapter). These studies primarily examined features of individual languages in Belize, rather than the multilingual practices involving alternation between languages within a

clause. Additionally, these were more descriptive than empirical (e.g. Le Page & Tabouret-Keller, 1985).

However, more recent studies by Balam (e.g., 2016a; 2016b; 2021), Balam and Parafita Couto (2019), and Fuller Medina (e.g., 2016; 2020b; 2021) have reported on naturalistic, corpus-based multilingual speech outcomes. For instance, Fuller Medina (2020b:8) conducted a comparative quantitative analysis on CS, focusing on different types of CS, based on naturalistic speech in Northern and Western Belize (for the analysis method, see Poplack & Meechan, 1998). She identified three main strategies of CS in these studies: (i) categorical integration of English-origin (including Kriol) nouns and verbs into Spanish grammar, (ii) alternation between languages in mid-sentence without violating the grammars of the respective languages, and (iii) use of English-origin (including English-based Kriol) discourse markers for pragmatic functions in Spanish. These observations correspond to Muysken's (2013) typology of CS strategies: Fuller Medina's (2020b) category (i) falls under 'insertions', category (ii) can be placed under 'alternation' and 'congruent lexicalization', and category (iii) is similar to 'backflagging', although these speakers are not heritage speakers.

Fuller Medina (2021) examined the Anglophone (English-origin) nouns and verbs inserted into Spanish-dominant clauses (with Spanish as the ML) and found that 66% (1739) of the code-switched cases were single item insertions, while 34% (893) were multiword fragments. Of these 1739 single items, 57% are nouns. This aligns with previous literature stating that nouns are the most regularly inserted parts of speech in the non-native language[5] (or, in this case, with Spanish as the *lingua franca*, Anglophone nouns). This percentage was followed by discourse markers (18%), verbs (13%), adjectives (6%), adverbs (3%), and others (3%), including conjunctions, kinship terms, numbers, and forms of address (such as Mr., Mrs., etc.). While Fuller Medina (2021) solely mentions that she extracted and categorized Anglophone speech segments from the Spanish-dominant clauses, the multilingual distribution of the Spanish-dominant versus Anglophone-dominant clauses remains unclear. Considering the prevalence of Spanish-

---

[5] As previously mentioned, reported in, e.g., Musken (2000) and Matras (2009).

dominant clauses in naturalistic speech, it suggests that Spanish as the ML is productive in Northern and Western Belizean multilingual communities.

2.3 Methods in CS research

In recent decades, numerous techniques have been applied to research CS phenomena using different approaches (i.e., linguistic, sociolinguistic, and psycholinguistic) for both qualitative and quantitative research (Jones, 2023). While the study of CS research originates in sociolinguistics, often through sociolinguistic interviews, the methodology has expanded to include various experimental techniques. One such technique, described in our previous section, is the comparative quantitative analysis carried out by Fuller Medina (2020a) on a dataset with naturalistic speech. Jones (2023) reviews the methods used in qualitative CS research and notes that complementary techniques and approaches are necessary to better understand different aspects of CS.

Gullberg et al. (2009) comprehensively explain several methods in CS research. One is the collection of a corpus of spontaneous speech. The most significant advantage of collecting spontaneous speech is that it provides the closest approximation to the community's most natural form of speech. However, collecting spontaneous speech is a challenging task. Gullberg et al. (2009) identify three disadvantages: (i) most bilingual or multilingual corpora are not publicly available, (ii) the process is costly and time-consuming, and (iii) all collected data must be transcribed, which may take months or years. Additionally, there is no conventionalized way to annotate and code bilingual or multilingual speech, which can be problematic for comparison purposes.

2.4 Building a multilingual corpus

To address these challenges, several (free) programs and transcription software are available to transcribe spontaneous speech data, each with slightly different purposes for various aspects of speech analysis. For instance, PRAAT is commonly used for analyzing phonetic aspects of unilingual or multilingual speech. The precision of the onset and clear visibility of the spectrogram (oscillogram) allows the annotator to

meticulously mark soundwaves and analyze them within the PRAAT software (Boersma & van Heuven, 2001; De Jong, Pacilly & Heeren, 2021; de Boer & Heeren, 2020).

Another well-known transcription program used in the transcription of bilingual speech is CHAT/CLAN. One of the few larger bilingual corpora publicly available, BangorTalk, also used this software for their transcripts.[6] A significant advantage of CHAT/CLAN is that multiple parts of information can be tagged in bilingual speech production, such as part of speech, glosses, and tags of multiple languages. Multilingual tags are essential in the analysis of bi- or multilingual speech. In addition to their transcripts, Carter, Broersma, Donnelly & Konopka (2018) developed an autoglosser for Welsh-English and Spanish-English speech. The highly accurate autoglosser output was then integrated back into their original CHAT/CLAN transcripts.

For the present study, the transcription software ELAN will be used.[7] The interface of ELAN is user-friendly, and a great advantage is that the video player and waveform are both visible while transcribing, which can be done via different modes. Due to this interface's setup, ELAN is nowadays widely used to annotate gestures and bimodal speech. It can also visualize a spectrogram such as PRAAT, although precise selection and analysis cannot be performed within the program. Multiple language tagging is possible; language tiers are created separately in ELAN instead of marking (tagging) the words in the same transcription tier or line as in CHAT/CLAN. This allows the content of each tier to be structured hierarchically with a parent tier. The audio of each speaker can be selected while the speech is annotated simultaneously (see Figure 1 for the ELAN interface). In ELAN, tags can be added in the form of a tier, with each tier created to carry information for each annotation.

---

[6] BangorTalk: BangorTalk bilingual conversational corpora

[7] Free to download on Download | The Language Archive.
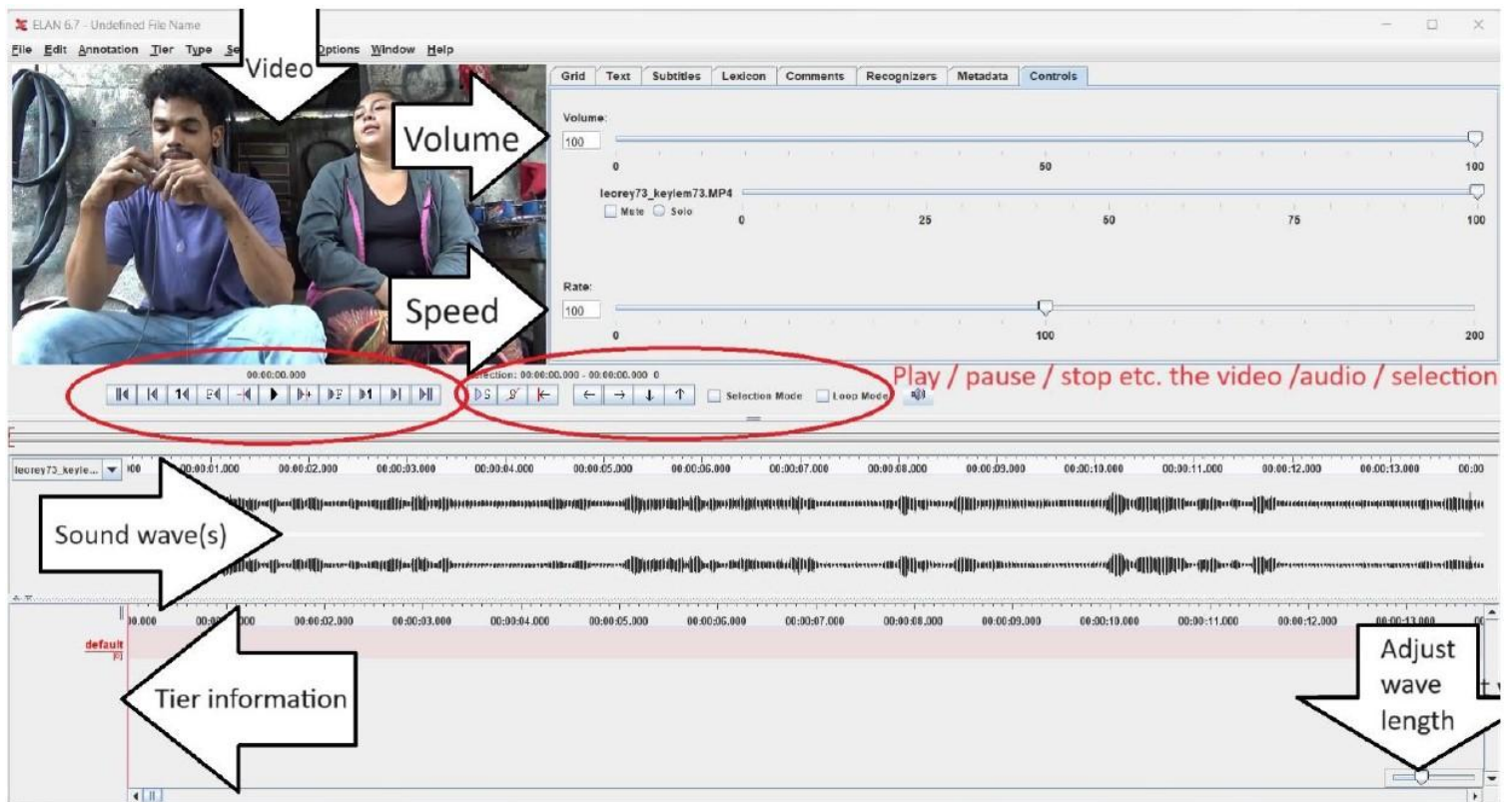
Figure 1. *ELAN interface.*[8]

2.4.1 Transcription conventions

As described earlier, transcribing a naturalistic corpus is a time-consuming matter (Gullberg et al.,2009). Spontaneous speech often lacks clear sentence boundaries and includes hesitations (e.g. '*s-s-s-o*') and repetition (e.g. '*well, well*'), fillers (e.g. '*uhm, uh, hmm*'), backchannels (e.g. as a response to the conversational partner '*of course, sure*'), tag sentences (e.g. '*you know?*'), half-words or even non-recognizable words or sentences. Additionally, other linguistic information relevant to discourse research, such as unfilled pauses as discourse markers, overall speech rate, and fluency, must be considered (Edwards & Lampert, 1993; De Jong, Pacilly & Heeren, 2021). Spontaneous speech also includes other sounds, such as laughter, other expressions of emotions, coughing, or throat clearing (i.a., Edwards & Lampert, 1993; Crible et al., 2017).

---

[8] Figure taken from the ELAN Coding Guide, created by the author of this study, see Appendix E for a link to the guide.

Unpredictable events can occur during the recording process: for example, if the camera operator accidentally creates a sound and participants react to it, or if someone spontaneously joins the conversation, turns on the radio, and the rest of the conversation becomes unclear. In other words, spontaneous speech is messy and somewhat unpredictable, both for the researcher recording the speech and for the transcriber observing the output. Numerous unforeseen events can occur during recording, and there is only so much the researcher can prepare for (e.g., good equipment and a quiet place to record the audio/video).

Conventions can be established when preparing the data for transcription and coding, but other issues will inevitably arise, as not every event in the data can be anticipated (e.g., a monolingual speaker joining a multilingual conversation). In such a large body of data, numerous research studies can be conducted, each requiring a slightly different method of transcription and coding based on the study's purposes. This challenge is further compounded if the recordings consist of multilingual or multimodal speech. For instance, fillers are language-specific (e.g., Spanish 'eh' versus Dutch 'uh', Crible et al., 2017). Determining what information should be included and how to organize those strings of speech and other sounds is crucial. Additionally, tagging the specific information needed to answer the research questions of the particular study is essential. Therefore, clear conventions need to be established prior to (and during) the transcription process.

In this chapter, we reviewed the types and strategies of CS, the MLF model and studies which applied this theoretical framework. Additionally, we tapped into CS research methods, focusing on multilingual corpus building. The following chapter will provide contextual information on the Northern Belizean participants of our study.

3. The multilingual context of Belize

The present study examines data from Orange Walk Town in *Northern Belize*. Therefore, we mainly focus on this area's history and current languages. The languages that we will focus on in this study are Spanish, Kriol, and English. We will elaborate on each of these languages and its particular linguistic aspects in Northern Belize. A more detailed exploration of language attitudes and the social context surrounding the primary languages is provided in Section 5 and 6 of this chapter.

Belize[9] - formerly known as British Honduras - is a relatively small country on the Central-American Caribbean coast, with 13.000 square kilometers of territory (about 260 km from north to south and 100 km from west to east). It is enclosed by Mexico (north), Guatemala (west and south), and the Caribbean Sea (east) and divided into six Districts, as illustrated in Figure 2. Belize has a vast ethnic and linguistic diversity (Escure, 2006), considering the total number of solely 400.000 inhabitants (The World Bank, 2024).

---

[9] According to Belizean legendary stories (and still a popular myth), even to be found in earlier history books, the name *Belize* was considered to be derived from a Scottish buccaneer called Wallace, who got shipwrecked on the Belize River in 1638 (or any other version close to this). In literature nowadays, however, we find that the name was most likely used by the British settlers, who referred to a minor waterway that the Mayans called 'Balis or 'Baliz'. This waterway by which the British could reach the main Belize River, was crucial to the British and their settlements. For this reason, references to it were productive in their written documentations. In modern Yucatec Maya, this means 'muddy' or 'muddy waters'. While the meaning of the 16th Century Mayan word 'Balis' cannot entirely be accounted for, most researchers agree on the fact that the name *Belize* derives from the 16th Century waterway called 'Balis' or 'Baliz' (Shoman, 2010; Bulmer-Thomas & Bulmer-Thomas, 2012).
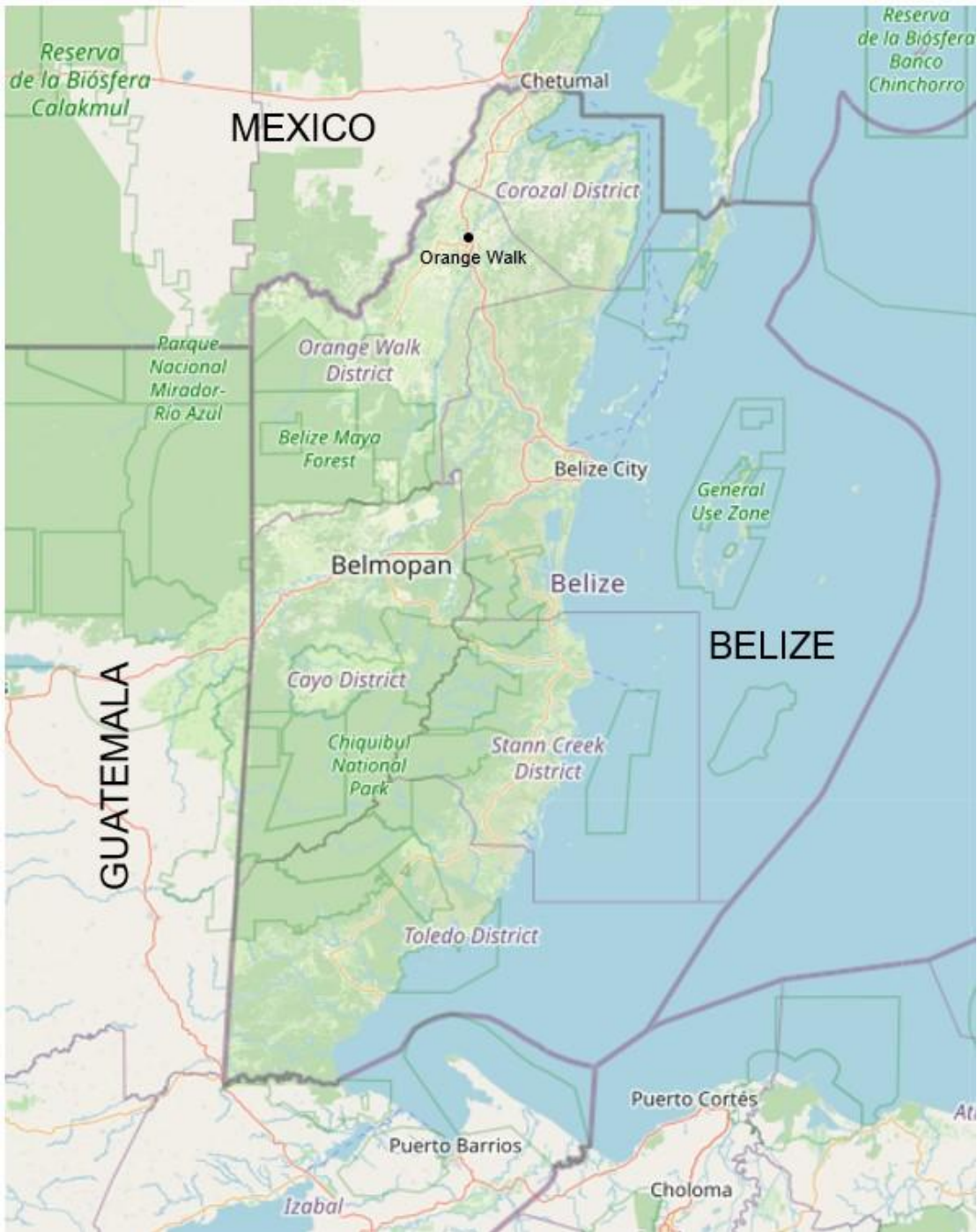
Figure 2. *Map of Belize and its six Districts: Corozal and Orange Walk to the north, Cayo, Stann Creek, and Belize in the center, and Toledo to the south. © OpenStreetMap Contributors, licensed under Open Database License, "ODbL" 1.0.*

3.1 Historical context of multilingual Belize

When Spanish settlers arrived in Belize in the 16th century, many of the once-flourishing Mayan civilizations had already declined significantly. The remaining Mayan populations primarily resided in remote areas, where Spanish influence was limited. For example, the Mayan settlement of Lamanai in Northern Belize actively resisted Spanish domination (Rushton, 2014). In the 17th century, British buccaneers arrived in Belizean territory, initially seeking to exploit the region's natural resources by harvesting and trading logwood (circa 1660–1910) and later mahogany (circa 1750–1945) along the Central American Caribbean coast. The Battle of St. George's Caye in 1798 marks the most decisive conflict between Spain and Britain, where the Spanish were defeated. After this, Belize officially became an English-speaking nation, and the Spanish colonial presence was diminished (Rushton, 2014; Fuller Medina, 2020a). In the centuries since the arrival of the British, the British privateers expanded their settlement along the coast. They began developing the area into an important international trade post, for which reason they also brought enslaved people from West Africa. In fact, the Caribbean coastal area was the only British outpost in Central America, as the rest of the region was dominated and governed by the Spanish (Escure, 2004) (see Figure 3).

Figure 3. *Locations of current Central American varieties of English (marked with a large black dot, Escure, 2004:518).*

Belize, self-governed since 1964 – gained its independence from Great Britain in 1981, after a long period of negotiating with neighboring Guatemala[10] (Holm, 1983; Straughan, 2004). The British colonial rule brought English as the national language, which is nowadays used in formal settings, including formal broadcasting, written

---

[10] The Guatemalan claim on Belizean territory (in the border area) remains unsolved until date, for which reason the bilateral relationships between Guatemala and Belize are complex. Guatemala has not officially recognized Belize's national status, thus, until this day, Belizeans fear annexation by Guatemala (Pérez & Sippola, 2021:296-297).

communication, and education (Pérez & Sippola, 2021:295). Nowadays, most Belizeans grow up speaking at least three languages (Escure, 1997:37; Pérez & Sippola, 2021). The three most common languages are English, Kriol, and Spanish, according to the Census 2022 (Statistical Institute of Belize, 2022). Kriol is reported to be the *lingua franca* of Belize (in general), and Spanish is the *lingua franca* of the Northern Districts of Belize (Orange Walk and Corozal) (Balam, 2016a). Despite this, previous studies also report on intense CS practices between multiple languages all across Belize (e.g., Balam, 2016a; Ravindranath Abtahian, 2017; Schneider, 2021; Bero, 2022).

3.2 Belizean Spanish

Today, Spanish is the predominant language in Northern Belize, shaped by two significant waves of Spanish-speaking immigration. The first wave occurred in the mid-19th century with the arrival of over 10,000 Mayans and Mestizos (individuals of mixed Indigenous and Spanish heritage) fleeing the Caste War on the Yucatán Peninsula. Many of these immigrants were bilingual in Yucatec and Spanish. While the exact trajectory of linguistic contact following this influx of non-English speakers remains unclear, it is likely that Yucatec initially served as the *lingua franca* in Northern Belize. Over subsequent centuries, a gradual linguistic shift occurred, transitioning from Yucatec to Spanish as the dominant language in these early Yucatec-Spanish communities (Cal & Fuller Medina, 2017; Fuller Medina, 2020a).

The second wave of immigration took place during the 1980s and 1990s, when approximately 30,000 to 40,000 predominantly Spanish-speaking individuals from Guatemala, El Salvador, and Honduras migrated to Belize. These movements were driven by civil wars in their countries of origin. As a result of this substantial influx, Spanish was officially recognized by the Belizean government in the 1990s. Although English remains the medium of instruction in schools, Spanish was incorporated into the curriculum as a foreign language (Schneider, 2021).

Despite this recognition, most contemporary immigrants settled in central districts rather than the northern regions of Belize, resulting in a dialectal distinction in Spanish usage across these areas (Balam, 2013, 2016a; Pérez & Sippola, 2021).

Contributing factors to this dialectal division include the waves of migration, the high number of Spanish speakers, and Belize's underdeveloped infrastructure until the mid-20th century. These dynamics led to the emergence of two distinct Spanish varieties in the northern and western regions; the Western Belizean Spanish (WBS) and the Northern Belizean Spanish (NBS)[11]. Notably, these varieties developed independently of formal prescriptive Spanish instruction, which only became part of the school system later. This absence of standardization likely contributed to the substantial dialectal and regional variation observed in Belizean Spanish (Fuller Medina, 2020a; Bero, 2022). Figure 4 illustrates the geographical distribution of Spanish dialects in Belize.[12]

Figure 4. *Dialect Map of Spanish in Belize (Cardona Ramírez, 2010:50). 1: NBS, 2: WBS*.

---

[11] Fuller Medina (2020b:21) remarks that, as stated by previous reports (cf. Cardona Ramírez, 2010:27,46), a large Spanish speaking community in the south of Belize possibly identifies a third Belizean regional variety of Spanish, with a similar history to that of the WBS and NBS varieties, although this needs further investigation.

[12] For an overview of Belizean Spanish linguistic features, see e.g. Fuller Medina (2020b:5).

3.3 Belizean Kriol

The terms *Bileez Kriol*, *Belize Kriol*, and *Belizean Kriol* (hereafter referred to as *Kriol*) all denote the same creole language spoken in Belize (e.g., Young, 1995; Belize Kriol Project, 1994, 1997; Greene, 1999). Kriol is an English-based creole, with 16th and 17th-century regional British English, likely Northern English and Scottish (see Holm, 1978), serving as its superstrate, primarily providing its lexicon. Other languages that influenced the creolization process of Kriol (substrates) are several African languages (Akan, Bantu, Yoruba, Twi, and Igbo (see Escure (2013), Fuller Medina (2020a)), Miskito (e.g., Holm, 1977, 1978), Garifuna (e.g., Escure & Schwegler, 2004), and minimally by Mayan languages (e.g., see Young (1995), for some lexical items). This section examines the social complexity surrounding Kriol, considering the identification process of being a Creole and speaking the Kriol language and the post-creole continuum. Attitudes toward Kriol will be explored in greater detail in the following sections.

3.3.1 Defining Belizean Kriol

To begin, it is essential to delineate the distinction between 'Creole and 'Kriol'. Although these terms are phonetically identical, the variation in spelling signifies their different meanings. 'Kriol' specifically refers to the language predominantly spoken in certain regions of Belize, primarily (although not necessarily) by individuals who identify as 'Creoles'; Belizeans who consider themselves Creoles [13] (Decker, 2013; Straughan, 2004). The rather complex identification process of being Creole is often intertwined, but not necessarily dependent on, speaking the Kriol language. Census 2010 data indicates that Kriol speakers (44,6%) outnumber those who identify as Creole (25,9%) (Fuller Medina, 2020).[14] This suggests that various (self-reported) ethnicities speak Kriol (Fuller Medina, 2020b:8).

---

[13] Following Decker (2013), we continue using the same terminology described for the rest of this study. With the term 'creole' or 'creole language' (without a capital C), we mean the general term for any creole language.

[14] This remark also counts for the more recent Census 2022 (Statistical Institute of Belize, 2022) data.

The term Creole originally referred to the ethnical mix between the West African slaves and their descendants, and the original Baymen, the British settlers along the Central American coast, who worked in the logwood industry. These enslaved West Africans were primarily brought via Jamaica, the administrative capital of the British settlements in the Caribbean, between 1720 and early 1800s, to help in the flourishing British wood industry. The descendants of these Africans often intermingled with other African descendants from other Anglophone parts of the Caribbean, such as Jamaica, Barbados, or other British establishments from the Caribbean coasts (such as the Bay Islands in Honduras or the Mosquito coast in Nicaragua). Later, the British and other Europeans also assimilated into the Creole population (Straughan, 2004). The identification of being a Creole nowadays depends sometimes on genetic ancestry but mainly shifts to cultural and linguistic behavior and a sense of belonging. For instance, an Afro-American descendant might not consider himself a Creole, while a Spanish-speaking immigrant who has adopted Kriol, identifies himself a Creole (and not a mestizo), as a matter of belonging, or 'being Belizean' (Decker, 2013:2).

That being the case, being Creole and speaking the Kriol language does not have a defined correlation, and both terms do not necessarily have delimited meanings. Furthermore, those who speak Kriol do not always define Kriol as a distinct language. As an English-based creole language, speakers might refer to it as 'bad English', and a 'corrupt', 'broken', or 'bastard' form of speech (Decker, 2013:4). In other words, they may consider the language as a substandard form of English, rather than recognizing it as a distinct language.

When examining speech data from a Belizean community, it is important to pay attention to language attitudes towards all the languages involved. This consideration is essential not only to shape the communities' ideologies and identities regarding the language(s), but because self-ratings of proficiency in Kriol or English might not be representative. According to Decker (2013:4), many Belizeans have insufficient access to a 'standard' variety of English (such as British English or American English) and do not recognize the extent to which Kriol differs from English varieties, thus considering Kriol as a form of English. Another explanation by Decker (2013:4) is that those exposed to these English

varieties may perceive Kriol as sufficiently similar to be considered a dialect, and they often label it a 'broken' or 'bad' form of English. Undoubtedly, these attitudes can cause difficulties in data collection (e.g., when creating a speaker profile or consulting a speaker to identify the language in question (Decker, 2013; Ravindranath Abtahian, 2017)). Acknowledging this sociolinguistic complexity, our study addresses these challenges and examines Belizeans' attitudes toward Spanish, Kriol, and English in greater detail in the following sections.

3.3.2 The post-creole continuum

Several theories exist on the origins of creole languages, with one of the key points being the debate on whether these languages should be considered unique in their development (see, e.g., *The Creole Debate,* McWorther (2018), or Mufwene (2001)). It is generally agreed that creoles languages originate from multiple languages within a colonial context. In the case of Belize, a diverse array of languages contributed to the emergence of Kriol. These included British English, Spanish, various indigenous languages (i.e., Mayan languages, Garifuna, Miskito), and African languages spoken by enslaved individuals brought to the Caribbean coast. Some of these languages may have had minimal influence, making it difficult to trace their impact on contemporary Kriol.

The continuous development of the creole language, which some scholars consider unique compared to other language's development (e.g., Decker, 2013)), has resulted in several varieties (lects) of Kriol. Factors such as changing attitudes towards Kriol, the frequency and locations of its use, and the modes of communication (spoken or written) all factors contribute to this ongoing development. When the creole language coexists with its superstrate language in a speech community, it may undergo a process of 'decreolization'. In this process, the creole language adapts linguistic features of the superstrate. As speakers incorporate these features, Kriol can develop into a range of spoken varieties (lects) (Decker, 2013). Today, Kriol serves as the *lingua franca* in Belize (except in Northern Belize), and its growing popularity and high identity value - especially since Belize's independence in 1981 – have led to the emergence of many lects. These lects range from basilectal varieties; typically unintelligible to Standard English speakers, to acrolect varieties; closely resembling standard forms of English, such as British or

American English (Velupillai, 2015). The numerous varieties between the basilect and acrolect are considered mesolects (see Table 1 for examples). Collectively, these lects form what is known as the post-creole continuum.

Table 1.

Examples of Kriol in the lectal continuum range from the basilects to the acrolects, with any other variety as a mesolect in between (Decker, 2013:5; adapted by Gómez Menjívar & Salmon, 2018:48).

Basilect Kriol	*Di flai dehn mi-di bait laas nait.*
Mesolect Kriol	*Di flies dem mi bitin las nite.*
	Dem flayz de baytin las nait.
	Di flayz-dem de waz baytin.
Acrolect Kriol	*Di mosquitos were bitin las night.*
Standard English	The mosquitos were biting last night.

The usage of each of these varieties differs across communities, individual speakers, and even within individuals in different social contexts. Understandably, when studying the production data of Kriol speakers, distinguishing between a standardized variety of English and acrolectal Kriol can be challenging. This is particularly true since Belize's official language is English, which coexists with Kriol. Therefore, it is essential to consider the post-creole continuum throughout our study to understand the distinctions between the Anglophone varieties of Belize.

3.4 Central American English and Belizean English variety

Given these complexities in the Kriol continuum, it becomes evident that speakers are not always aware of whether they are using Kriol or a standardized form of Belizean English, the official language of Belize (referred to as 'English' in this study). These

speakers might be unable to identify the differences between English and Kriol (acro)lects (Decker, 2013; Balam, 2016a). This ambiguity in language identification presents challenges not only for the speakers themselves, but also for the researchers working in this linguistic landscape. Besides, the literature review reveals some uncertainty regarding the current usage and categorization of this official English.

Geographically, the official English of Belize falls under the span of Central American English (CAE) varieties. The collection of English in this region spreads all along the Caribbean coast, from Belize to Panama (see Figure 5).
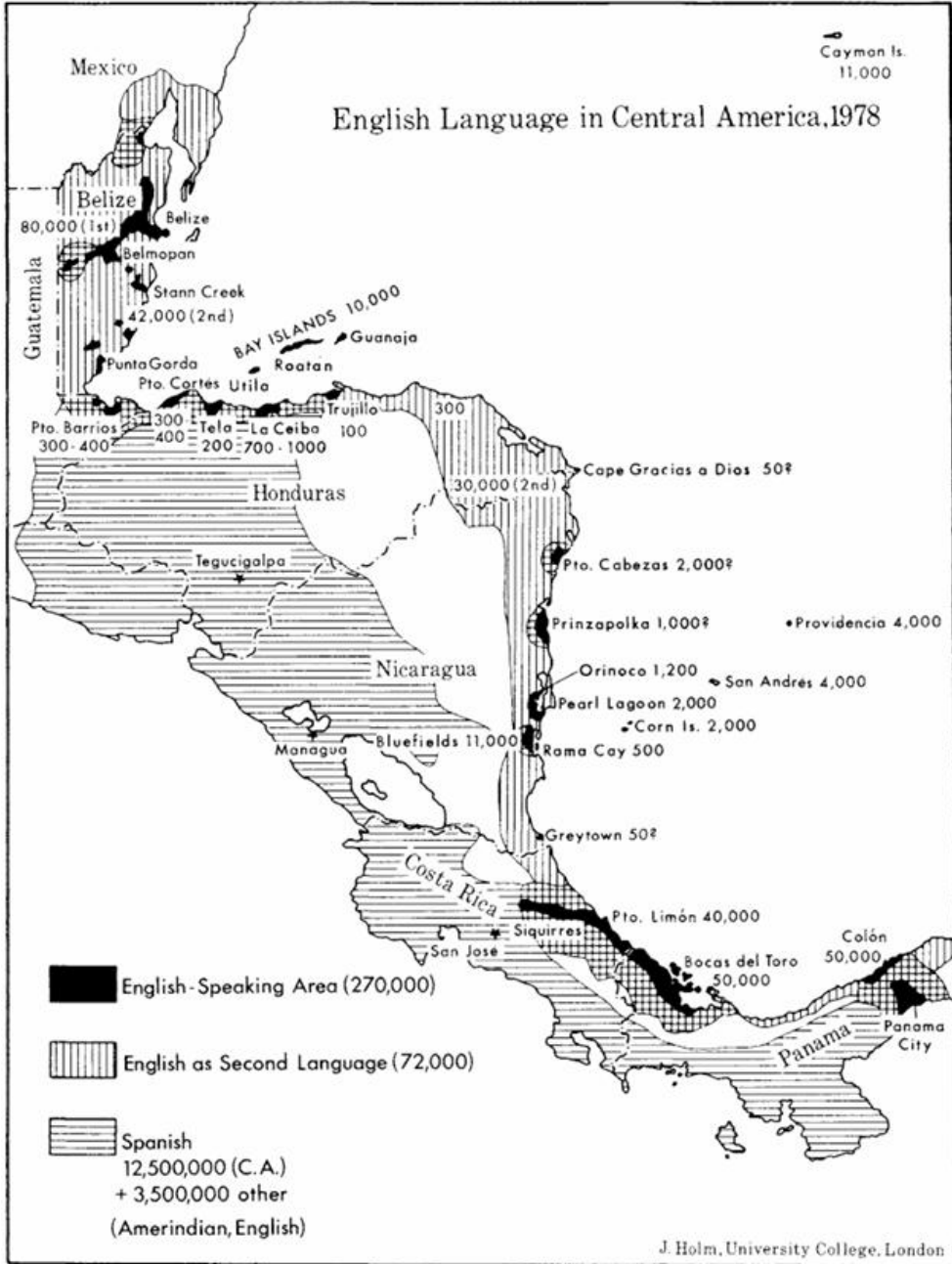
Figure 5. *English Language in Central America in 1978 (Holm, 1983:8).*

Worth to consider, as Holm (1983) points out, is the fact that these varieties mostly refer to the *spoken* English in Central America, while the written form usually adheres to Standard English of Great Britain or the United States. Occasions where this 'Standard English' is spoken are generally specific, such as in a ceremonial sermon of a Nicaraguan bishop in Belize (Holm, 1983:11). Many historical events and factors in the English speech communities, such as contact with Spanish, led to the widespread variations of CAE today. Consequently, CAE encompasses a large variety of spoken English. However, Holm's (1983) description of uniform aspects of CAE often reflects features of Kriol rather than English (e.g., in terms of plural endings, loans, or calques from Spanish or particularities in tense-mood-aspect systems).

This suggests that grammatically, the official English in Belize should not be categorized as one of the CEA varieties. This overlap complicates a precise definition of what constitutes the official English of Belize and how it is used today. Over 30 years later, Balam (2016a) reports that English is primarily used in written, and not in spoken form in Northern Belize. Although the usage of English appears to have evolved from 'none' spoken to 'almost never' spoken, the question remains as to what extent English is currently spoken in 2025. Given the rapid changes in globalization, modern technology and the ease of being exposed to other international varieties of English, it is worth considering an increase in the presence of spoken English today.

Therefore, for this study, we consulted a local school teacher in Orange Walk Town, Arturo Acosta,[15] to understand the current usage of English. He reported that English, as taught at school, is productively spoken inside and outside the classroom, while Kriol, Spanish, or other languages are not permitted during class. However, he reports that students frequently engage in CS outside the classroom in Kriol, English, and Spanish. Given that Belize was reigned initially by the British Crown, Arturo mentioned

[15] We are grateful for Arturo Costa for helping us by sharing his observations. He currently teaches at Muffles Junior College (Muffles Junior College), the school in Orange Walk Town, where most participants in our dataset study at the time of data collection. Please note that we do not intend to make definitive claims about the current presence of spoken English based solely on Arturo's observations. Instead, we primarily use this source of information to gain a general understanding of the current linguistic scope of official English usage in the area.

that there are no specific policies or protocols for English written or spoken in class, except that it must adhere to Standard English grammar rules. While study materials are in American English, both Standard American English and Standard British English orthographies are acceptable. Arturo also noted that, due to media access nowadays, students are more likely to adopt an American English accent, while a British accent is rarely heard and would sound odd to him. He noticed that students "*coming from the villages*" (near Orange Walk Town) are more likely to speak Spanish at home, or have a greater command of Spanish in general than the students raised in Orange Walk Town. This, in contrast with 'previous times' in which Spanish was the home language of more students. Arturo clarifies this with a personal anecdote: He raised his children primarily in English due to modernity and access to the language, a trend he sees among other educated parents as well. Consequently, his children have a lower proficiency in Spanish, making it difficult for them to communicate with their Guatemalan grandmother, who only speaks Spanish. He now feels he 'did wrong' and believes his children would need to work hard at school to achieve the necessary command of Spanish to ease communication.

In other words, it can be cautiously assumed that Belizean official English is no longer confined to written or strictly formal settings; rather, it appears to have gradually been incorporated into the spoken language of the community over the past decades.

3.5 Language attitudes in Northern Belize

While English is on the television, in the newspaper, in any governmental or formal situation, and at school, Spanish is taught (in English) at school as part of the curriculum since the 1990s (Gómez Menjívar & Salmon, 2018). Kriol is in larger parts of the country the most dominantly spoken language, while Spanish is the *linga franca* in Northern Belize. How these languages are preferred over national English might be explained by factors such as the speakers' language attitudes towards each of these languages.

After the Caste War in the second half of the 1800s, Northern Belize is dominated by Spanish speakers. Most probably, few English or Kriol speakers lived in the area by that time. The first language contact between Spanish and English/Kriol likely started with the

logwood or *chicle* industry in the first decades of the 1900s (Balam, 2013; Bero, 2022). Since halfway through the 19th Century, infrastructure has greatly improved across Belize and its borders with neighboring countries. This also impacted and increased both international and national migrations, and with this, more Spanish varieties were added. More English and Kriol speakers also came to live in these Northern Districts. Additionally, schooling started to be taught in English to the Mayans/mestizos of the area, and the rate of bilingual or multilingual speakers increased. These modern developments impacted the indigenous Mayan cultures and languages, and Mayan languages became uncommon, as English and Spanish were learned (Balam, 2013; Bero, 2022).

Given these shifts in the linguistic landscape that this area had to endure over the past 150 years, it is remarkable that Spanish remained the dominant language throughout the entire timeframe. Especially in the context of an official English-spoken country, despite the improved infrastructure and relatively close dominant English/Kriol-speaking Districts. Interestingly, this does not necessarily mean that the Spanish always had an overt prestige status in these northern communities. In fact, due to the many varieties of Spanish in a somewhat isolated form that developed over the years in the NBS and WBS varieties, that same Spanish had been reported to be 'broken Spanish' or 'Creole Spanish' (Le Page & Tabouret-Keller, 1985). This is due to the consciousness of the Belizean varieties being deviant from more standard Caribbean Spanish. A decade before this, Koenig (1975) found that *half* of the Spanish speakers expressed a negative attitude towards another language (in the Corozal District). In 80% of the cases, this negative language attitude was towards Kriol. Even more remarkably, half of the Kriol speakers had a negative attitude towards their own language. Thus, in the 70s and 80s, there was a specific ranking on attitudes towards Spanish and Kriol, with Kriol being the language of lowest prestige. As explained in the previous section, attitudes toward Kriol greatly improved due to language promotion and planning, especially from independence onwards. However, recent studies on Spanish attitudes sometimes still reflect the language views that were brought up in the 80s (Balam, 2013; Balam 2017; Bero, 2022).

Le Page (1992) (c.f. Balam, 2013) describes that the attitudes towards Spanish became even more negative when Guatemala put its claim on Belizean territory[16].

Balam (2013) finds in his study on language attitudes with younger participants (age 15-25, *n=42*) that most have a rather negative attitude towards the Spanish they speak. All participants agreed that NBS differs from Standard Spanish and that NBS is a 'lesser' variety than Standard Spanish. According to the participants, the latter is the 'proper Spanish', 'pure Spanish', 'real Spanish', 'Spanish Spanish', and so on. At the same time, the participants still preferred to hear the NBS over the Standard Spanish in their speaking environment. In general, male participants were neutral on the question of whether Standard Spanish should receive the same status as Standard English (taught in school), and female participants had a slightly stronger opinion on this matter: they were more in favor of Standard Spanish being in the same official status as English.

Furthermore, relatively little is known about language attitudes towards official English. In Balam's (2013:258) paper, one male participant (recorded in Orange Walk Town) expressed that he would find it 'irritating' if people spoke Standard Spanish or Standard English to him. On this topic, Balam (2013:258) also highlights that in Koenig's (1975:111) study, several students also disliked the use of Standard English instead of Kriol. Nevertheless, due to modern times and better access to a Standard English variety though media, this may shift. Students may also prepare for or seek out, i.e., an educational trajectory or joining the emigrated family members in the relatively proximate U.S.A. (Straughan, 2004; Daugaard-Hansen, 2009). This is possibly a strong motivation to adapt their English to Standard American English (Bero, 2022).

3.6 CS attitudes in Northern Belize

In a multi-ethnic and multilingual country such as Belize, CS is ubiquitous (Balam, 2016b; Fuller Medina, 2020a; Schneider, 2021; Bero, 2022). We should consider that the language attitudes described in the previous sections reflect attitudes focusing on these

---

[16] And as mentioned before, this debate is still ongoing. Keeping in mind that this possibly plays a role in the current attitudes towards NBS, WBS or Belizean Spanish in general.

languages used in isolation, in a unilingual preferred environment, and not in a multilingual environment. We also point out that these language attitudes may have been influenced by structural, historical inputs on language use, shaped by, e.g., political motivations, such as the formation of a nation. Also, speakers of these languages are aware of standardized forms around them; they may even be familiar with speaking these standardized forms (e.g., they compare the NBS with Standard Spanish or Kriol with a Standardized form of English). This may also influence their attitudes to what they could consider 'broken forms'. These points considered, attitudes may differ when these languages are mixed.

In the case of Belize, attitudes toward CS practices are generally positive or neutral and not negative in the few studies that have been carried out on multilingual practices (Bonner, 2001; Ravindranath, 2009; Balam, 2013, 2017; Bero, 2022). However, these attitudes were recorded in specific towns across the country (Orange Walk, Corozal, Hopkins, and Dangriga), and not across the entire country. Each of these studies had specific research goals. For instance, Balam (2013) investigated attitudes among 15-25 years-old adolescents towards NBS, standard Spanish, and CS attitudes in Orange Walk Town, and Balam (2017) reported on Spanish and CS attitudes of teachers' perceptions of student's Spanish and CS attitudes in the classroom in Northern Belize (it was in continuation of his Balam (2013) study). Bonner (2001) and Ravindranath (2009) studied the language use (and shift) in Garifuna-Kriol communities in Hopkins and Dangriga (in the Coastal area, not in Northern Belizean Districts).

Recent work that has been carried out on CS attitudes in Northern Belize specifically are those by Balam (2013) and Balam (2017). Generally, he found that CS is labeled as a prestigious practice, and using solely unilingual varieties is avoided. Bero (2022) conducted his study in the Northern Belizean multilingual environment of Corozal Town. Even though Bero (2022:51) did not investigate CS attitudes specifically, he reports that from the content of his sociolinguistic interviews, he got the impression that people included multilingual discourse as a form of social behavior. He speculates that individual differences in use and attitudes probably relate to lifestyle and individual experience.

In this chapter, we described the linguistic situation of Belize, both historical and contemporary, including its current languages. We also reviewed the language and CS attitudes throughout the past decades on these languages. In sum, we understand that Kriol is a creole language in an ongoing post-creole continuum, with acrolects potentially challenging to separate from Standard English, Belize's official (formal) language. The Spanish spoken in Northern Belize falls under a particular variety (NBS) with distinct grammatical features compared to other varieties in Belize or Central America. Due to this, language attitudes are somewhat divided: on the one hand, NBS is the *lingua franca*, so speakers prefer speaking this Spanish over Standardized Spanish or English and Kriol. On the other hand, speakers are aware that this form of Spanish differs from the Standard, and some consider this a 'lesser' form. Language attitudes on Kriol are nowadays overtly positive, in contrast with attitudes before the independence, and are mainly associated with complex intertwined identities and ideologies, especially from ideas originating from the '70s and '80s. Kriol seems to have gained the status of being a prestigious language in Belize in general.

4. Research questions

After examining the Matrix Language Frame (MLF) theoretical approach to CS practices, the linguistic landscape of Belize, and the social status of Spanish, Kriol, and English in the Northern region, we propose the following research questions for this study:

Research Question 1. *What is the distribution of unilingual as opposed to multilingual speech in Northern Belize, and is there a preference for the Matrix Language (ML) in code-switched speech?*

With the answer to this question, we aim to characterize the extent of multilingualism, CS practices, and language preferences in Orange Walk Town, Belize. Additionally, we seek to explore social factors influencing these CS practices, hence we pose the following question:

Research Question 2. *To what extent is there a relation between CS attitudes, proficiency, the languages used in participants' social networks, and the proportional outcomes of unilingual and multilingual speech?*

Based on the literature review, we hypothesize the answers to our research questions as follows:

Hypothesis 1. We hypothesize that CS practices are prevalent in Northern Belize, as supported by previous studies (e.g, Balam, 2014; Fuller Medina, 2021). Consequently, we expect a significant proportion of multilingual speech compared to unilingual speech in the conversational data of our participants. Furthermore, given that Spanish serves as the *lingua franca* in this community (Balam, 2016a), we anticipate a preference for Spanish unilingual clauses over English or Kriol unilingual clauses. Although literature has shown that ML preference in multilingual speech is not always evident in each community, (e.g., Carter et al., 2011), we expect any observed ML preference to be explained by underlying extralinguistic factors.

The complex multilingual landscape of Northern Belize results from its diverse population and colonial past, influenced by languages such as Spanish, Kriol, and English (Balam, 2014). The prominence of Spanish as the *lingua franca* is rooted in the historical migration patterns and social structures (Balam, 2016a). Research indicates that CS practices are a common feature in Northern Belizean speech communities. Studies by Balam (2014) and Fuller Medina (2021) have documented the widespread use of CS in daily communication in Northern Belize specifically. Carter et al. (2011) reported on three communities with different outcomes on the choice of the ML. The use of naturalistic conversational data ensures that our findings accurately reflect the authentic speech patterns of the participants (Gullberg et al., 2009).

Hypothesis 2. Numerous factors may influence the intensity of unilingual and multilingual language preferences, such as age, gender, proficiency, location, individual language attitudes, CS attitudes, and sociolinguistic contexts (e.g., Parafita Couto et al., 2014). Parafita Couto et al. (2014) reported in their study on naturalistic data that the choice of the ML could be related to social identity, social networks and possibly also education. We anticipate the emergence of a specific community standard for ML choice in unilingual and possibly in multilingual clauses due to sociolinguistic factors. This community-wide preference in ML has been previously identified by e.g., Blokzijl et al. (2017) and Carter et al. (2011), and explained by factors such as individual language attitudes, CS attitudes, proficiency levels and language use in the social network of speakers. We hypothesize that setting a homogeneous group and controlling for the variables age, gender, language proficiency, and residence (in this case, Orange Walk Town), will yield similar outcomes in unilingual and multilingual speech across the community. If variations are found, we expect them to be partially explained by individual differences in proficiency levels, CS attitudes, and the language use in social networks of the individual participants. Based on our literature review, we understand that attitudes toward the Northern Belizean Spanish (NBS) variant are neutral or positive (e.g., Balam, 2013), and attitudes toward Kriol are overly positive, especially in contrast to pre-independence attitudes (e.g., Gómez Menjívar & Salmon, 2018; Abtahian Ravindranath, 2017). Attitudes toward English are unclear; however, English appears to have been

incorporated into spoken speech in Northern Belize. In the few studies that focused on CS attitudes within this community, we understand that these are generally positive (Balam, 2013; 2017). The languages used within participants' social networks may influence their speech patterns. For example, individuals who frequently interact with Spanish speakers may produce more Spanish clauses. By controlling for age, gender, proficiency, and other variables, we aim to isolate the specific factors influencing unilingual and multilingual speech patterns. This approach allows for a more nuanced understanding of the relationship between sociolinguistic factors and language use.

5. Methodology

As part of the larger research project Crossing Language Borders[17], 50 recordings of multilingual spontaneous speech data were collected in the Orange Walk Town area in Northern Belize in November and December of 2023. Each recording features a 15 to 20-minute conversation between two participants, resulting in approximately 15 hours of spontaneous speech data. The participants were not directed to discuss specific topics; they were encouraged to converse freely in front of a camera operated by one of the project researchers.[18] Participants were not required to use any particular language, nor were they instructed on how to use their languages. Thus, the corpus data can be considered a spontaneous speech production.

However, it needs to be clarified that the participants were aware that the study focused on code-switching, as indicated in the information sheet provided at the beginning of the data collection process (see Appendix A, see Appendix B for the informed consent form). Additionally, most participants completed an extensive background questionnaire before the conversation (some afterwards). This questionnaire was administered by the researcher, who recorded the answers on a computer while the participants responded. The questionnaire included questions about the participants' personal details, such as age and gender, as well as linguistic information, such as language proficiency, age of onset for each language, code-switching attitudes, and the

---

[17] This project is collaborative initiative among international scholars across several linguistic research institutes (i.e., ACLC, LUCL, USCD), funded by the NWO. The research group in this project is interested in the study of the human language capacity in a variety of ecologies involving multilingualism, language contact, and code-switching/mixing, and in how the interaction between the multilingual mind and these ecologies leads to linguistic changes. See also Crossing Language Borders - ACLC - University of Amsterdam (uva.nl) for more information on the project.

[18] The main researcher in this data collection was Renzo Ego Aguirre Santa Cruz, he was not part of the Belizean speech community. As part of the Crossing Language Borders project, Renzo had a local mentee that he trained to collect linguistic data, Emmanuel Arcia. Emmanuel also operated the videorecorder numerous times in the dataset and he collected the background information of the participants via the background questionnaire. Emmanuel was part of the speech community of Orange Walk Town. For this reason, there is a chance to find differences in the output files when Renzo or Emmanuel operated the camera (Observer's Paradox, Labov, 1972).

languages used within their social network (see Appendix E for a link to the complete questionnaire). As a result, most participants were well aware of the project's objectives, which involve research on mixing languages and attitudes toward CS.

The relationships between the conversational partners in the video recordings varied and included couples, family members, friends, classmates, neighbors, and colleagues. This study's multimethod comparative approach connects sociolinguistic factors to grammatical outcomes (Munarriz & Parafita Couto, 2014).

5.1 Choosing a dataset for the current study

At the outset of this study, the 50 video recordings had not yet been transcribed. Our goal was to identify a dataset within those recordings, comprising 100 participants, that was as homogeneous as possible. We carefully selected conversations based on several factors: age, gender, languages spoken and their proficiency, and the town/area in Belize where they were raised and/or currently reside. By considering these factors, we aimed to achieve a representative outcome in our results.

The size of our dataset is based on the transcription of 18 participants in 9 video recordings, with a total duration of approximately 140 minutes. The selection process of this dataset will be described in the following sections, with each section addressing a specific factor that influence multilingual speech production. First, we will examine the frequencies and distribution of these factors in the entire dataset (*n=100*), including age, gender, spoken languages, and linguistic area. Next, we will illustrate the dataset (*n*=18) that will be utilized in current study. Finally, we will present the CS attitudes of the 18 selected participants.

5.1.1 Age (*n*=100 and *n*=18)

The full dataset (n=100) encompasses participants ranging in age from 18 to 85 (see Graph 1). Approximately half of this age range is concentrated in individuals under the age of 30 (see Table 2).

Graph 1. *Age range and its proportion in the entire dataset (n=100).*

Participants' age range and proportion
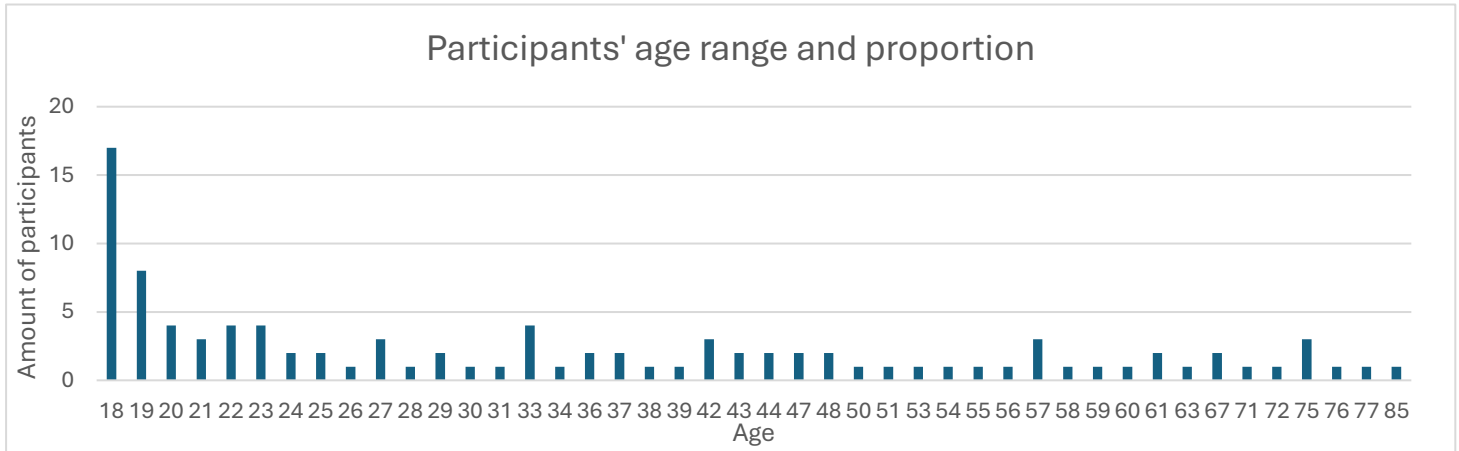
Table 2.

Frequency of participants for each age ≤ 30 and the (cumulative) percentages (%) within the full dataset (n=100).

Age	Frequency	Percentage of Ntotal (%)	Cumulative Percentage of Ntotal (%)
18	17	17	17
19	8	8	25
20	4	4	29
21	3	3	32
22	4	4	36
23	4	4	40
24	2	2	42
25	2	2	44
26	1	1	45
27	3	3	48
28	1	1	49
29	2	2	51
30	1	1	52

Table 3.

Frequency of participants for each age and the (cumulative) percentages (%) within the sample dataset (n=18).

Age	Frequency	Percentage of Ntotal (%)	Cumulative Percentage of Ntotal (%)
18	11	61,1	61,1
19	3	16,7	77,8
23	1	5,6	83,3
24	1	5,6	88,9
25	1	5,6	94,4
76	1	5,6	100

As shown in Table 2, approximately half of the participants in the entire dataset (*n*=100) are in the age range of 18 to 29 years old, while the other half is distributed across ages 30 to 85. For the current study, we selected 18 speakers within the age range of 18 to 29, preferably as close in age as possible, representing 40% of the entire dataset. By choosing participants from the same age group, we aim to exclude the possibility that (multilingual) speech production differs due to age. Above, Table 3 illustrates the age distribution within our study's dataset. As an exception to the rest of the files, we included one recording featuring a 76-year-old grandfather (*elamog*) in conversation with his 18-year-old grandson (*adrtun*).

5.1.2 Gender (*n*=100 and *n*=18)

Another factor that may influence multilingual speech production, is the gender of the speaker. All participants had a conversational partner, who sometimes had the same gender, either female-female (FF) or male-male (MM), while in other cases, the gender of the participant pairs was mixed, female-male (FM). Table 4 shows the distribution of the participant gender pairs across the 50 recordings. Table 5 provides an overview of the gender pair distribution in our sample set.

Table 4.

Overview of the gender pair distribuition among the participants in the full dataset (n=100).

Overview Gender	Number of recordings	% of Ntotal
FF	20	40
FM	15	30
MM	15	30

Table 5.

Overview of gender pair distribution of the sample dataset used in the current study (n=18).

Overview Gender	*Number of recordings*	% of Ntotal
FF	3	33,3
FM	2	22,2
MM	4	44,4

5.1.3 Multilingual participants (*n*=100 and *n*=18)

Participants rated their proficiency in each language and were asked to distinguish their rating for *speaking* and *understanding* it. Almost all of the participants (withthree exceptions) in the entire dataset claimed that they *speak*, or at least *understand*, the following three languages: English (En), Spanish (Sp), and Kriol (Kr). Some participants also speak and/or understand a language from the Mayan language family: Mopan (MoM), Qeqchi (QeqM), or Yucatec (YucM). One speaker is also familiar with German (Ger). Table 6 provides an overview of the languages spoken in the entire dataset.

Table 6.

Overview of the language spoken by the participants Ntotal=100.

	n=	% of Ntotal
En/Sp/Kr	90	90
Ger/En/Sp/Kr	2	2
MoM/En/Sp/Kr	1	1
QeqM/En/Sp/Kr	2	2
YucM/En/Sp/Kr	2	2
YucM/En/Sp	3	3

Table 7.

Overview of the language spoken by the participants of the dataset of the current study Ntotal=18

	n=	% of Ntotal
En/Sp/Kr	17	94,4
Ger/En/Sp/Kr	0	0
MoM/En/Sp/Kr	0	0
QeqM/En/Sp/Kr	0	0
YucM/En/Sp/Kr	1	5,6
YucM/En/Sp	0	0

As shown in Table 6, except for three participants (3%), all others speak English, Spanish, and Kriol. 90% of all participants speak and/or understand *only* these three languages. The remaining 7% also speak either German or a language from the Mayan family. Table 7 provides an overview of the language distribution of this study's dataset.

5.1.4 Hometown of the speakers (*n*=100 and *n*=18)

Besides the age, gender, language background, and proficiency of each participant, all participants were asked where they currently live, where they previously lived, and for how long. Most of the 100 participants resided in Orange Walk Town at the time of the data collection or in one of the nearby villages, such as Trial Farm, Nuevo San Juan, and Yo Creek. In the past, a few participants (or their parents) had immigrated to Belize from neighboring countries such as Mexico, Guatemala, El Salvador, or Honduras. However, most participants had previously lived in Orange Walk Town, its surrounding villages, or other cities in Belize, such as Belmopan or Belize City. For the dataset of the current study, we selected participants who had lived most of their lives in or around Orange Walk Town. Two participants were from the nearby villages of Trial Farm and Nuevo San Juan.

5.2 Overview of the dataset of this study (*n*=18)

In summary, based on all the previously described information, the dataset of this study consists of the following participants: they are all proficient in En/Sp/Kr and all from Orange Walk (OW) or the surrounding villages. The nine recordings contain three pairs of female speakers, two pairs of mixed speakers, and four pairs of male speakers. They are all 18-25 years old, except for one 76-year-old speaker. Participants were selected based on age, gender, language proficiency, and their current residence, with a preference for those raised in the area. This set of participants is detailed in Table 8.

Table 8. *Participants of the dataset used for the current study.*

File name	Participant ID	Age	Gender	Education	Language	Prof. En	Prof. Sp	Prof. Kr	Current region	Previous region	Occupation
keylemleorey121223ow	keylem	23	F	U. Secondary	En/Sp/Kr	4/4	4/4	4/4	OW	Orange Walk (2000-2023)	Head cook
keylemleorey121223ow	leorey	25	M	Primary	En/Sp/Kr	4/4	4/4	4/4	OW	Orange Walk (1998-2023)	Bicycle mechanic
adrtunelamog171223ow	adrtun	18	M	L. Secondary	En/Sp/Kr	4/4	4/4	4/4	OW	Aldea Paraíso (Corozal) (2005-2021)	Student
adrtunelamog171223ow	elamog	76	M	H. Education	En/Sp/Kr/YucM	4/4	4/4	3/4	OW	Aldea San Lázaro (1947-1962) Belize City (1962-1968) Orange Walk (1968-1972) Belize City (1972-1974) Orange Walk (1974-2023)	Retired (educator
rongonzuruk241123ow	rongon	18	M	L. Secondary	En/Sp/Kr	4/4	2/3	3/3	OW	Orange Walk (2005-2023)	Student
rongonzuruk241123ow	zuruk	18	M	L. Secondary	En/Sp/Kr	4/4	3/3	3/4	OW	Aldea Douglas (2005-2012) Orange Walk (2012-2023)	Electrician
carchaemaort171223ow	carcha	24	F	Primary	En/Sp/Kr	4/4	4/4	2/3	OW	Orange Walk (1999-2009) Aldea Benque Viejo (Cayo) (2009-2018) Orange Walk (2018-2023)	Church helper
carchaemaort171223ow	emaort	19	M	L. Secondary	En/Sp/Kr	4/4	4/4	2/2	OW	El Loco Tío, Honduras (2004-2008) Orange Walk (2008-2023)	Student
daycasjaiben221123ow	daycas	18	F	L. Secondary	En/Sp/Kr	4/4	3/4	2/3	OW	Orange Walk (2005-2023)	Student
daycasjaiben221123ow	jaiben	18	F	L. Secondary	En/Sp/Kr	4/4	2/2	4/4	Trial Farm	Aldea Trial Farm (2005-2023)	Student
glekanvivnov291123ow	glekan	18	F	L. Secondary	En/Sp/Kr	4/4	4/4	4/4	OW	Orange Walk (2005-2023)	Student
glekanvivnov291123ow	vivnov	19	F	L. Secondary	En/Sp/Kr	4/4	2/3	4/4	OW	Orange Walk (2004-2023)	Student
evedeljosnah161123ow	evedel	18	M	L. Secondary	En/Sp/Kr	3/4	4/4	2/3	San Felipe	Aldea San Felipe (2005-2023)	Student
evedeljosnah161123ow	josnah	18	M	L. Secondary	En/Sp/Kr	4/4	2/3	4/4	OW	Orange Walk (2005-2023)	Student
emmarcshecar181223np	emmarc	18	M	U. Secondary	En/Sp/Kr	4/4	4/4	2/4	OW	Dangriga Town Stann Creek District (2005-2009) San Estevan Village (2009-2019) Orange Walk Town (2019-2023)	Student/call center agent
emmarcshecar181223np	shecar	18	M	L. Secondary	En/Sp/Kr	4/4	4/4	2/1	OW	Aldea San Estevan (2005-2019) Orange Walk (2019-2023)	Student / gas station assistant
jacblakiamai211123ow	jacbla	19	F	L. Secondary	En/Sp/Kr	4/4	4/4	2/3	Nuevo San Juan	Aldea Nuevo San Juan (2004-2023)	Student
jacblakiamai211123ow	kiamai	18	F	L. Secondary	En/Sp/Kr	4/4	3/3	2/3	OW	Orange Walk (2005-2023)	Student

5.2.1 Participant's background (*n*=18)

All participants lived in Orange Walk Town (OW) or within 2 kilometers of it (Trial Farm) at the time of research, with the exception of one participant who lived in Nuevo San Juan, 19 kilometers north of OW. Participants self-rated their proficiency in each language, distinguishing between *speaking* and *understanding*. The rating scale used was as follows: '4' for 'excellent', '3' for 'good', '2' for 'fair', and '1' for 'poor'.[19]

All participants, except one (*evedel*), rated themselves as 'excellent' in English in both *speaking* and *understanding*. Four out of eighteen participants rated themselves as less proficient in Spanish, with at least a 'fair' rating in *speaking* and mostly a 'good' rating in *understanding*. The most significant variation in language proficiency ratings was found in Kriol. Seven out of eighteen participants rated their Kriol *speaking* proficiency as 'fair', whereas sixteen out of eighteen participants rated their Kriol *understanding* proficiency as 'good' or 'excellent'. The two exceptions (*emaort* and *shecar*) rated their Kriol *understanding* proficiency as 'fair' or 'poor'.

The educational background of all participants ranges from primary school to lower secondary education (L. Secondary), upper secondary education (U. Secondary), and higher secondary education (H. Secondary). As most participants were students, this educational background may still be ongoing.

---

[19] The proficiency rating in the column for each language in Table 8, consists of two numbers (e.g., 3/4). The first number (e.g., 3) is for the self-rated proficiency for *speaking* the language. The second number (e.g., 4) is for the self-rated proficiency for *understanding* the language.

Table 9.

Background information on the age of acquisition (AoA) and CS attitudes.

Participant ID	AoA En	AoA Sp	AoA Kr	How often CS?	In which context?	"What do you think about this kind of practices?"
keylem	4	3	13	always	close family/relatives/couple/coworkers/neighbors	*"It's useful because everybody does it."*
leorey	3	3	7	often	close family/relatives/couple/coworkers/neighbors	*"It's good and useful, because we are a melting pot of cultures so we can communicate fluently."*
adrtun	3	3	7	sometimes	close family/relatives/couple/classmates/neighbors	*"It's unique because two languages come together as one and it's an art."*
elamog	5	3	16	always	close family/relatives/couple/coworkers/neighbors	*"It's wonderful, because we can communicate with many people easily."*
rongon	3	5	7	often	close family/relatives/couple/classmates/neighbors	*"I think it's a unique concept that a country only like Belize can pull off."*
zuruk	8	3	11	often	close family/relatives/coworkers/classmates/neighbors	*"It's unprofessional because you may blurt out a bad thing in front of your superiors leaving you in a bad light."*
carcha	3	3	10	sometimes	close family/relatives/coworkers/neighbors	*"It's good because you can learn maybe words that we don't know."*
emaort	4	3	10	sometimes	close family/relatives/coworkers/classmates/neighbors	*"It's wrong, I think every language is different and they should be separated. In my case, I want to speak Spanish in a proper way."*
daycas	5	3	13	always	close family/relatives/classmates/neighbors	*"It's interesting and useful because I use another language to complete my ideas."*
jaiben	3	3	3	often	close family/relatives/classmates/neighbors	*"It's good, beautiful and useful."*
glekan	5	4	4	always	close family/relatives/classmates/neighbors	*"This kind of practice allows us to connect with each other through language. I can talk to all the people in my family, at school, and even un my future work environment due to the many languages spoken."*
vivnov	3	10	6	often	close family/relatives/couple/classmates/neighbors	*"It's good because that helps us how to communicate with one another."*
evedel	10	3	14	often	relatives/couple/classmates/neighbors	*"It's necessary, because you can complete your ideas with other languages."*
josnah	4	3	3	often	close family/relatives/couple/classmates/neighbors	*"It's unique, but not the right thing to do, it can be problematic when we are required to speak just in one language."*
emmarc	5	3	16	always	close family/relatives/couple/coworkers/classmates/neighbors	*"It is a unique practice and is awesome."*
shecar	3	3	14	always	close family/relatives/coworkers/classmates/neighbors	*"It's normal, so I don't care."*
jacbla	4	3	13	always	close family/relatives/couple/classmates/neighbors	*"It's normal."*
kiamai	3	3	7	often	close family/relatives/classmates/neighbors	*"It's ok because it's helpful when I want to complete an idea, I can use another language."*

5.2.2 CS ratings and attitudes (*n*=18)

In the background questionnaire, each speaker was asked: "Does it happen that you use several languages within a single conversation?". If they answered affirmatively, they rated the frequency of such occurences. In Table 9, these answers range from 'never', 'rarely', 'sometimes', 'often', to 'always'. As observed in the column 'How often CS?', most participants declared to code-switch 'often' or 'always'. Three participants reported that they 'sometimes' code-switch, and none stated that they 'never' or 'rarely' engage in CS. This indicates that CS is a common practive among these participants. Overall, the opinions about CS are positive (see all reactions in the right column of Table 9). Interestingly, only three participants expressed somewhat negative views, despite regularly practicing CS ('unprofessional' (*zuruk*), and 'wrong' (*emaort*), 'unique, problematic'(josnah)). Better yet, for most participants CS is so ordinary that some even described it as 'normal', and most mentioned all the contexts suggested in the questionnaire. An exception is *evedel*, who stated he does not code-switch with his family (or in his home, as indicated in another question/answer not mentioned here). Given that he learned English at age 10 and Kriol at age 14, he was probably raised in Spanish and learned the other languages later. Although remarkably, he still considers CS 'necessary'.

Two other important details about the age of acquisition mentioned in this table. First, none of the participants considered the ages '0' or '1' as their age of acquisition for any languages they learned. For this reason, we consider the age '3' as the starting point for their first languages. This implies that some participants have two or three first languages. In some cases, the ages of acquisition are sufficiently different to indicate they are clearly raised in one language. For instance, participants *vivnov*, *zuruk*, and *evedel*, have a gap of 3, 5, or 7 years, respectively, between their first and second acquired languages. However, the exact gaps may be slightly uncertain, as they must have started speaking their first language before age three. Another important detail is that Kriol is often acquired at a later stage in life. On average, the ages of acquisition for the languages are: English= 4.3 years old, Spanish= 3.6 years old, and Kriol= 9.7 years old. Considering the

dataset, the default age of language acquisition ('from birth') is set at 3 years old. Therefore, Kriol is not typically the first language of the participants in this dataset.

In this chapter, we have carefully selected a dataset with spontaneous speech production from 18 participants, based on their age, gender, and self-rated proficiency in English, Spanish, and Kriol. Additionally, participants were selected from the same hometown or surrounding villages. With this dataset, we explore their multilingual spontaneous speech production through a corpus-based analysis. The process of building and analyzing this corpus will be described in the following chapter: Transcription.

6. Transcription

With a team of three transcribers (Renzo, Kiki, and Emma), we needed to establish transcription conventions to ensure the reliability of our study. As previously mentioned, there is no standardized method for annotating and coding bilingual or multilingual speech due to varying research purposes (Gullberg et al., 2009; Jones, 2023). However, some publicly available bilingual corpora provide insights into data transcription and the associated challenges. Deuchar, Davies, Herring, Parafita Couto, and Carter (2014) describe the process of the methods used to design and build the three corpora in BangorTalk. They included information on participant recruitment, recording procedures, strategies to avoid the Observer's Paradox (Labov, 1972), and the use of CHAT/CLAN for corpus construction. Given the involvement of multiple transcribers over the course of several years, the authors decided to use Turnitin, a commercial plagiarism detection service, to test the reliability of the transcripts. They compared transcripts of the same audio fragment with Turnitin and generated quantitative measures (percentages) of transcript concordance. Additionally, Deuchar, Donnelly, and Webb-Davies (2018) provided an in-depth account of the transcription process for one of these corpora, the Siarad corpus, highlighting the specific challenges faced in transcribing bilingual speech.

6.1 Transcription conventions

In alignment with the research questions of this study, the transcription using ELAN requires the following information: (i) determination of the linguality (unilingual/multilingual) of each clause, (ii) determination of the language(s) in all clauses, and (iii) determination of the Matrix Language (ML) of the clause. For further investigation of code-switched cases, we will continue coding in Excel. After carefully reviewing the literature on transcription and coding of (bilingual) naturalistic speech and considering the research questions of our study, we established transcription conventions, as illustrated in Appendix C.

We aim to investigate CS patterns and relate the outcomes to the MLF model in this study. This model accounts for and makes predictions about code-switches on a clausal level.

Therefore, transcription of the data needs to be on a clausal level, rather than in turns, full sentences, or other divisions of speech. Additional information such as repetitions, hesitations, and fillers should be registered to identify outliers or individual cases. This can provide information on the speaker's CS fluency, for instance. We also considered the readability of the transcription. For example, marking a word with 'bus@eng@spa' (English and Spanish language tagging, as in CHAT/CLAN), reduces readability, especially when the entire clause is filled with other tags. To maintain simplicity, we did not add tags to our clauses, and we documented any remarks in a separate file. We decided to annotate the speech, mark the linguality of the clauses in ELAN, and code the rest in Excel. This was a more practical consideration, as it is not feasible to work on a single file with multiple persons in ELAN. However, by exporting all transcripts from ELAN to Excel, we could code each column with multiple persons in a shared spreadsheet.

6.2 Transcription process

For the transcription process, we maintained an Excel log file to annotate anything remarkable, unrecognizable, or unclear across each recording, as illustrated in Appendix D (see Appendix E for a link to the complete transcription issues file). Additionally, to validate our transcripts, a fourth transcriber (Simon) reviewed 3 out of 9 recordings for inconsistencies, unclear parts, and other remarks. These comments were also logged in (see also Appendix E). We tracked the status of each transcription in our Belize Transcription Progress File (for an example see Appendix D, and for a link to the complete Transcription Progress File see Appendix E).

6.3 Transcription tiers in ELAN

For this study's spontaneous speech corpus data, we created two tiers: the transcription tier and the linguality tier (see Figure 6).
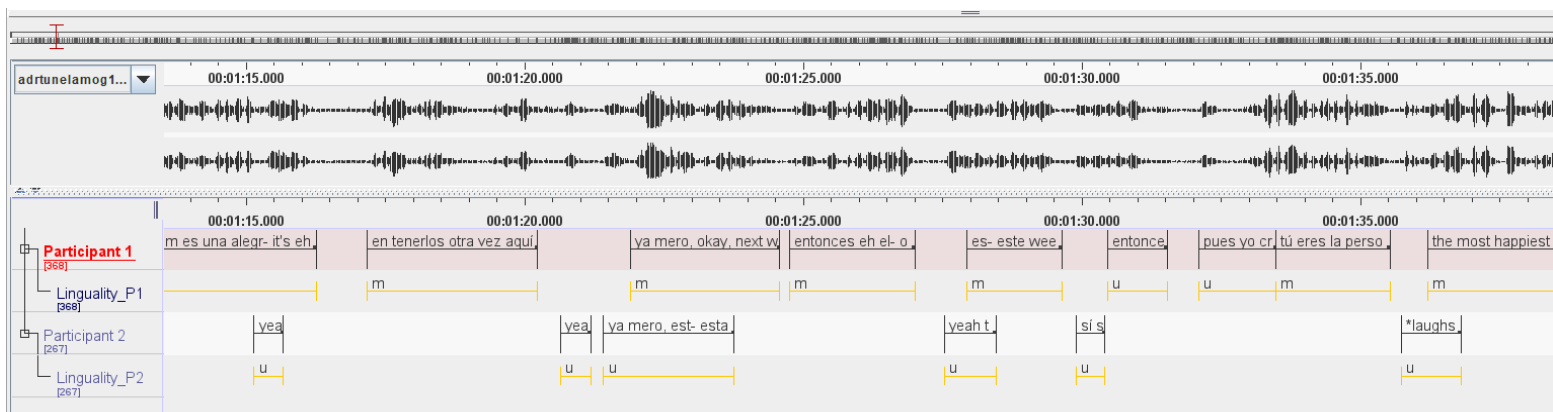
Figure 6. *A close-up of several transcription tiers in ELAN. The tiers are hierarchically structured and contain the content per participant, per clause, and the linguality per clause.*

6.4 Further considerations

Transcribing in Kriol orthography already inherently distinguishes whether the speech is English or Kriol. Despite occasional grammatical differences, it usually remains unclear which of the languages should be annotated. Furthermore, English words can appear in Kriol grammar. None of the transcribers are familiar with Kriol. This means that the transcribers require frequent usage of the Kriol dictionary. There will be a higher chance of inconsistencies when doing this. Therefore, for this study, we transcribe the words in Standard American English as they are being said. In continuation, we use an auditory impressionistic analysis when identifying the Kriol or English language (Fant & Tatham, 1975). We further explain this in 7.2.3.

Given the social and historical context of the multilingual community in Northern Belize, we may expect ethnical connotations to each of the community languages. Furthermore, we may expect a high frequency of CS in the form of backflagging. We may find differences in proficiencies due to the historical context. Therefore, while the participants in our study are not heritage language speakers, they may have learned one or two languages at a later age (such as Standard English at school). Fuller Medina (2021) found many discourse markers, which may be ethnically or socially connotated. We will include discourse markers in our coding as well. We understand that these code-switches are peripheral to the clause and not part of it. Therefore, we mark these clauses as unilingual in (around) clauses where no other type of CS occurs.

7. Coding & analyses

7.1 Coding and analyses in Excel

We further coded the nine transcribed and partly coded video recordings in Excel. We imported the ELAN tiers of participants p1 and p2 and linguality p1 and p2 into separate columns in Excel.

7.2 Coding per transcript

We coded four columns for each transcript: clause/other, linguality, language, and matrix language (see Figure 7). In the following paragraphs, we elaborate on our work methods for each.

	file_name	participant_1_emaort	participant_2_carcha	linguality_	languag	claus	matrix_
2	carchaemaort171223ow.eaf	good morning *name*		u	e	o	
3	carchaemaort171223ow.eaf	cómo estás?		u	s	c	s
4	carchaemaort171223ow.eaf		morning, bien, muy bien, y tú?	u	se	o	
5	carchaemaort171223ow.eaf	bien aquí, solo en banquitas		u	s	o	
6	carchaemaort171223ow.eaf		qué cuentas de UNCLEAR,	f	su	c	s
7	carchaemaort171223ow.eaf		como te está yendo?	u	s	c	s
8	carchaemaort171223ow.eaf	bueno, quería ir ayer a ver al parade,		m	se	c	s
9	carchaemaort171223ow.eaf	pero no pude pertener work		m	se	c	s

Figure 7. *An impression of 9 rows of coded transcript. UNCLEAR means the transcriber could not hear the speech clearly (e.g., due to background noise).*

7.2.1 Clause/other

As described in the previous chapter, naturalistic speech can contain many other speech segments than clauses. Therefore, apart from the clause, we separated the 'other' speech segments in a new line. Each main clause and subordinate clause was also separated into new lines. Speech segments containing an UNCLEAR were coded as 'f' (flexible).

Examples: unilingual 'other' (15), unilingual 'clause' (16), multilingual 'other' (17), multilingual main 'clause' (18), multilingual subordinate 'clause' (19).

15. 'two hours before yeah'

> (participant_1_emaort, carchaemaort171223ow.eaf)

16. 'como   te   est-á   yendo?'
    how   2SG.DAT   be-3SG   go.PROG
    "how is it going?"

> (participant_2_carcha, carchaemaort171223ow.eaf)

17. '**con el**   new   **líder**'
    with DEF.SG.M   leader
    "with the new leader"

> (participant_1_elamog, adrtunelamog171223ow.eaf)

18. '... **pero en este**   season **que está**   **hech-a**   decorated'/
    but   in   DEM.SG.M   that   be.3SG.AUX.PRS make-PTCP.SG
    "... but in this season that it was decorated"

> (participant_2_carcha,carchaemaort171223ow.eaf)

19. / 'that I'm studying it   **ahor-ita**'
                 now-DIM.SG.F
    "that I'm studing it right now"

> (participant_2_adrtun ,adrtunelamog171223ow.eaf)

7.2.2 Linguality: unilingual/multilingual

We included discourse markers, fillers, and tags along the same lines as the clause, as those are part of the speech. As we annotated all speech produced, we encountered several cases where the clause was started in one language, corrected, and translated into another. In these cases, we counted the clauses 'unilingual' as well. When the clause was coded 'multilingual' due to a real code-switch, the clause remained 'multilingual' when discourse markers were also in different languages within the same line. Speech

segments containing an UNCLEAR and solely having one language were coded as 'f' (flexible) (see also Figure 7). Speech segments containing an UNCLEAR and two languages were coded as 'multilingual'.

Examples: 'unilingual' (20), 'unilingual' with discourse marker 'alright' (21), 'unilingual' with tag sentence (22), 'unilingual' with translation (23), 'multilingual' (24), 'multilingual' with discourse marker 'pues' (25).

20. 'it would be better for him to wait for at at at at at that time, or or?'

(participant_2_leorey, keylemleorey121223ow.eaf)

21. '**ajá entonces ehm**   alright, **ya**   **est-amos**'

aha then   already   be-1PL.PRS

"aha, then uhm alright, we are done"

(participant_1_elamog, adrtunelamog171223ow.eaf)

22. '**él**   **tiene**   **para**   **mí**,   you know'

3SG   have.3SG.PRS for   1SG.DAT

"he has (it) for me, you know"

(participant_1_glekan, glekanvivnov291123ow)

23. '**antes   de que** before it gained independence'

before   of that

"before it before it gained independence"

(participant_2_adrtun, adrtunelamog171223ow.eaf)

67

24. '**te**   **llev-é**   **un**   burger **y**   **un**   coke'

   2SG.DAT   take-1SG.PST   INDF.SG.F   and   INDF.SG.F

   "I brought you a burger and a coke"

<div align="right">(participant_2_shecar, emmarcshecar181223np)</div>

25. '**y**   **él**   **hiz-o**   postpone **pues** **su**   **su**   **ehm** concert'

   and 3SG   do-3SG.AUX.PST   well   POSS.3SG   POSS.3SG

   "and he postponed, well, his his uhm concert"

<div align="right">(participant_2_kiamai, jacblakiamai211123ow)</div>

7.2.3 Language: English/Spanish/Kriol

In this tier, we marked the language for each clause (or 'other' speech). In Chapter 2, we explained that, due to the creole continuum, acrolect forms of Kriol are close to English and cannot easily be separated. This was challenging in our dataset when we marked the language for our speech segments. We solved this in three different ways.

First, by doing an impressionistic auditory analysis (Fant & Tatham, 1975) for each segment. We carefully considered whether the anglophone speech sounded more like Kriol or English. As outsiders of this community, they were differentiated between 'clearly understandable' for English and 'less understandable' or 'hard to understand' for Kriol. In practice, this mostly meant we had an impression of the speech (per speaker) of the entire video recording and sometimes parts of the video recording. It was more challenging on a clausal level. However, secondly, when we noticed that the grammar was clearly Kriol and not English, or vice versa, we marked the language according to that. Lastly, we looked at each participant's self-rated Kriol and English proficiency. English *speaking* was always marked 'excellent', so when Kriol *speaking* was marked '2' (fair) or '1' (poor), it was likely that this person would not engage much in Kriol's speech. Especially if the conversational partner self-rated their Kriol *speaking* similarly. In these cases, we marked the entire anglophone speech of these participants as 'English' unless an auditory impression (in this case 'less understandable' or 'hard to understand') or

grammar (on a clausal level) would tell us otherwise. Speech segments containing an UNCLEAR and solely having one language were coded as 'u' (unknown) (see also Figure 7).

Examples: Kriol based on auditory impression (26), Kriol based on grammar (27 and 28).

26. '*she move on go before Christmas*'

    "she comes back before Christmas"

<div align="right">(participant 2_jaiben, daycasjaiben221123ow.eaf)</div>

27. '*well basically yes because of if he reach tomorrow at eight*'

<div align="right">(participant_1_keylem, keylemleorey121223ow.eaf)</div>

28. '*we're gonna no have food / for come back *laughs*, right*?'

    "we're not going to have food / for when she comes back, right?"

<div align="right">(participant 2_jaiben, daycasjaiben221123ow.eaf)</div>

7.2.4 Matrix language: English/Spanish/Kriol

For the coding of the ML, we solely concentrated on the clauses, not 'other' segmented speech. As described in section 2.2, we pointed out that the ML can be determined by the language of the system morphemes, such as inflectional morphemes, determiners, or pronouns. For this reason, we identified the ML by verb inflection.

8. Results

8.1 General CS patterns

In this chapter, we present the findings of our corpus-based analysis of multilingual language patterns that emerge in the speech of the 9 video recordings (around 140 min).

8.1.1 Proportions of unilingual/multilingual speech

Our data of 18 video-recorded participants show a total production of 5337 segments of speech. 717 of these segments (13%) were multilingual, as illustrated in Table 10.

Table 10.

A total number of produced unilingual and multilingual speech by the 18 participants.

	amount	percentage		amount	percentage
total_production	**5337**	**100%**	**total_production**	**5337**	**100%**
total_unilingual	4279	80%	total_clauses	4010	75%
total_multilingual	717	13%	total_other	1155	22%
total_flexible	341	6%	total_flexible	172	3%

The numbers on the left are the overall numbers of the entire speech production (based on our 'linguality' coding tier), including the minor numbers of tokens, such as backchannels 'hmm' or 'yeah sure.' In 341 segments (total_flexible, 6%), we were either (i) not being able to hear the speaker completely (annotated solely as 'UNCLEAR') or (ii) partly not being able to hear the speaker (e.g., 'yeah, I live around twenty-three UNCLEAR'[20]). On the right side of Table 10, we split the total speech production (5337 segments) differently. Namely, we based this on our 'clause/other' coding tier. Here, we can observe that the speech of this dataset consists of 4010 clauses and 1155 'other' segments (examples described in our previous chapter, section 7.2.1). There were 172

---

[20] However, this was coded as a clause, as this speech segment contains a subject and predicate.

cases (3%) in which we could not define whether the speech segment was a clause or 'other'. Combining these two tiers, linguality and clause/other, we can observe the total amount of unilingual and multilingual clauses/other, as shown in Table 11.

Table 11.

Total number of clauses and other production by the 18 participants.

	amount	percentage		amount	percentage
total_clauses	**4010**	**100%**	**total_other**	**1155**	**100%**
total_unilingual_clauses	3214	80%	total_unilingual_other	1070	93%
total_multilingual_clauses	632	16%	total_multilingual_other	85	7%
total_flexible_clauses	164	4%			

In Table 11, we find 632 multilingual clauses out of a total of 4010 (16%) clauses, of which 3214 were unilingual (80%). In 164 cases (4%), we could not define the linguality of the clauses, yet we could determine the speech segment as a clause.[21] In the distribution of the linguality relative to 'other' segments, we see that only 85 cases (7%) were multilingual. Comparing this with the amount of multilingual clauses, the number is less than half. This may be due to the length of those segments, as those are usually smaller than multilingual clauses, meaning there is a smaller segment in which a code-switch can occur. Worth mentioning, in total, we found 84 cases in which unilingual produced segments (clauses or other) included discourse markers in the other language. In 52 cases, these discourse markers were Anglophone, peripheral to a unilingual Spanish speech segment. Vice versa, we solely found 21 cases. The rest were speech segments such as 'yeah, sí'. In 34 cases, these discourse markers were found at the start of the speech segment; in 26 cases, this was at the end; in 13 cases, those occurred at multiple places, such as in Example 29. In the rest of the cases, these discourse markers were found in the middle of the speech segment.

---

[21] As in the case of the example in the previous section, 'yeah I live around twenty-three UNCLEAR'. This example contains a subject and a predicate, therefore it is a clause.

29. aha, **sí**, yeah they've **ehm**, footbreaks, **y-, sí**

"aha, *yes*, yeah, they've *uhm* footbreaks, *and-*, *yes*"

<div align="right">(participant_2_leorey in keylemleorey121223ow.eaf)</div>

In 4 cases, tag sentences of the other language occurred, in an otherwise unilingual speech segment (3 times Anglophone 'you know' and 1 time Spanish 'no sé' ('I don't know')). There were 12 cases where a translation occurred in the same speech segment, eight from Spanish to Anglophone and three from Anglophone to Spanish. In those 11 cases, the speech segment started in one language, was translated into the other language, and continued for the rest in the latter. The one exception was as follows: 'correct correct correct **correcto** correct', in which the translation seems more a discourse strategy (giving emphasis perhaps) rather than a correction in speech (as in the other cases).

8.1.2 Proportions per participant

We will continue displaying these numbers of speech production for each participant now. This is shown in the following two pages in Table 12. The conversational partners are next to one another (except for daycas and jaiben). Besides the total amount of speech production and the amount clauses, we also added the (matrix) languages of the unilingual/multilingual clauses. We will further elaborate on this in our next paragraph.

Table 12.

Amounts and percentages of speech production of our dataset (n=18), to be continued on the next page.

	keylem		leorey		elamog		adrtun		zuruk		rongon		emaort		carcha		daycas	
	number	%	number	%	number	%	number	%	number	%	number	%	number	%	number	%	number	%
total_production	**288**	**100%**	**187**	**100%**	**364**	**100%**	**274**	**100%**	**304**	**100%**	**269**	**100%**	**235**	**100%**	**335**	**100%**	**358**	**100%**
total_unilingual	278	97%	179	96%	222	61%	229	84%	285	94%	214	80%	158	67%	259	77%	336	94%
total_multilingual	10	3%	8	4%	142	39%	45	16%	6	2%	29	11%	63	27%	58	17%	21	6%
total_flexible	0	0%	0	0%	0	0%	0	0%	13	4%	26	10%	14	6%	18	5%	1	0%
total_production	**288**	**100%**	**187**	**100%**	**364**	**100%**	**274**	**100%**	**304**	**100%**	**269**	**100%**	**235**	**100%**	**335**	**100%**	**358**	**100%**
total_clauses	274	95%	146	78%	254	70%	176	64%	183	60%	174	65%	172	73%	272	81%	281	78%
total_other	14	5%	41	22%	110	30%	98	36%	114	38%	77	29%	61	26%	56	17%	77	22%
total_flexible	0	0%	0	0%	0	0%	0	0%	7	2%	18	7%	2	1%	7	2%	0	0%
total_clauses	**274**	**100%**	**146**	**100%**	**254**	**100%**	**176**	**100%**	**183**	**100%**	**174**	**100%**	**172**	**100%**	**272**	**100%**	**281**	**100%**
total_unilingual	264	96%	140	96%	123	48%	134	76%	175	96%	144	83%	104	60%	209	77%	261	93%
total_multilingual	10	4%	6	4%	131	52%	42	24%	2	1%	22	13%	57	33%	52	19%	19	7%
total_flexible	0	0%	0	0%	0	0%	0	0%	6	3%	8	5%	11	6%	11	4%	1	0%
total_unilingual_clauses	**264**	**100%**	**140**	**100%**	**123**	**100%**	**134**	**100%**	**175**	**100%**	**144**	**100%**	**104**	**100%**	**209**	**100%**	**261**	**100%**
total_english_kriol	209	79%	104	74%	43	35%	68	51%	171	98%	99	69%	27	26%	86	41%	247	95%
total_spanish	55	21%	36	26%	80	65%	66	49%	4	2%	45	31%	77	74%	123	59%	14	5%
total_multilingual_clauses	**10**	**100%**	**6**	**100%**	**131**	**100%**	**42**	**100%**	**2**	**100%**	**22**	**100%**	**57**	**100%**	**52**	**100%**	**19**	**100%**
total_matrix_english_or_kriol	4	40%	5	83%	12	9%	18	43%	1	50%	4	18%	2	4%	4	8%	11	58%
total_matrix_spanish	6	60%	1	17%	119	91%	24	57%	1	50%	18	82%	55	96%	48	92%	8	42%

	jaiben		glekan		vivnov		evedel		josnah		emmarc		shecar		jacbla		kiamai	
	number	%	number	%	number	%	number	%	number	%	number	%	number	%	number	%	number	%
total_production	**312**	**100%**	**316**	**100%**	**200**	**100%**	**230**	**100%**	**175**	**100%**	**558**	**100%**	**407**	**100%**	**222**	**100%**	**303**	**100%**
total_unilingual	306	98%	293	93%	167	84%	211	92%	152	87%	364	65%	254	62%	160	72%	212	70%
total_multilingual	6	2%	19	6%	30	15%	5	2%	19	11%	96	17%	57	14%	42	19%	61	20%
total_flexible	0	0%	4	1%	3	2%	14	6%	4	2%	98	18%	96	24%	20	9%	30	10%
total_production	**312**	**100%**	**316**	**100%**	**200**	**100%**	**230**	**100%**	**175**	**100%**	**558**	**100%**	**407**	**100%**	**222**	**100%**	**303**	**100%**
total_clauses	250	80%	301	95%	175	88%	204	89%	143	82%	415	74%	190	47%	165	74%	235	78%
total_other	62	20%	15	5%	24	12%	24	10%	32	18%	97	17%	144	35%	49	22%	60	20%
total_flexible	0	0%	0	0%	1	1%	2	1%	0	0%	46	8%	73	18%	8	4%	8	3%
total_clauses	**250**	**100%**	**301**	**100%**	**175**	**100%**	**204**	**100%**	**143**	**100%**	**415**	**100%**	**190**	**100%**	**165**	**100%**	**235**	**100%**
total_unilingual	244	98%	279	93%	147	84%	188	92%	122	85%	284	68%	120	63%	120	73%	156	66%
total_multilingual	6	2%	18	6%	27	15%	5	2%	18	13%	79	19%	48	25%	33	20%	57	24%
total_flexible	0	0%	4	1%	1	1%	11	5%	3	2%	52	13%	22	12%	12	7%	22	9%
total_unilingual_clauses	**244**	**100%**	**279**	**100%**	**147**	**100%**	**188**	**100%**	**122**	**100%**	**284**	**100%**	**120**	**100%**	**120**	**100%**	**156**	**100%**
total_english_kriol	241	99%	242	87%	85	58%	184	98%	85	70%	8	3%	10	8%	6	5%	6	4%
total_spanish	3	1%	37	13%	62	42%	4	2%	37	30%	276	97%	110	92%	114	95%	150	96%
total_multilingual_clauses	**6**	**100%**	**18**	**100%**	**27**	**100%**	**5**	**100%**	**18**	**100%**	**79**	**100%**	**48**	**100%**	**33**	**100%**	**57**	**100%**
total_matrix_english_or_kriol	2	33%	8	44%	6	22%	2	40%	0	0%	2	3%	1	2%	0	0%	1	2%
total_matrix_spanish	4	67%	10	56%	21	78%	3	60%	18	100%	77	97%	47	98%	33	100%	56	98%

74

8.2 CS patterns and the MLF model

Besides the 'linguality' and 'clause/other' tiers, we also coded for 'language'. This, together with the 'matrix language' tier, informs us how Kriol, English, and Spanish proportions are distributed. It will also tell us which language shapes multilingual clauses in terms of morphosyntactic content. Table 12 shows these language distributions for each individual in the 'total unilingual clauses' and 'total multilingual clauses' rows. Although we distinguished between Kriol and English in our transcripts, in the manner described in section 7.2.3, we combined English and Kriol in this table.[22] Therefore, a clarification on the amount of English/Kriol ML in unilingual and multilingual clauses is needed. Hence, the amount of English/Kriol in unilingual clauses can indicate that (i) these clauses are indeed unilingual, in either English or Kriol, (ii) these clauses are for the greater part multilingual, due to CS between English and Kriol on a clausal level, and (iii), the combination of unilingual (in either English or Kriol) and multilingual clauses. Nevertheless, the amount of English/Kriol in multilingual clauses continues to be either Kriol or English as an ML. There is no change in the manner in which we perceive and compare the ML for multilingual clauses. A low number of unilingual Spanish clauses, in combination with the participant's self-rated report on high CS engagement, will possibly give us estimates of how these English/Kriol unilingual clauses are composed. In such a case, it may be likely that a greater part of the unilingual English/Kriol clauses are actually multilingual. Perhaps self-rated proficiency is another factor that determines the English/Kriol distribution in unilingual clauses. If the speaker's English or Kriol has a low proficiency rate, it is presumable that most - if not all - unilingual clauses are unilingual. We will further evaluate and elaborate on these speculations in section 8.4.

---

[22] It was a confronting task for us, outsiders, to understand and transcribe Kriol, a language none of us transcribers listened to before. Consequently, distinguishing between these languages was challenging, yet it was manageable for the greater parts within the files. This, due to the three ways to go about it, as described in our section 7.2.3. However, for the purpose of this study, we needed this information on a clausal level, to see if any code-switches occurred between English and Kriol. We noticed in the process that this was a rather ambitious task. For this reason, instead, we combined the languages into the greater 'anglophone' category.

Table 13 provides an overview of the languages of the total amount of clauses of all 18 participants. The data suggest a slight preference for unilingual English/Kriol clauses (total of 1921 clauses, 60%) versus unilingual Spanish clauses (total of 1293 clauses, 40%). Considering that English and Kriol are combined, some unilingual clauses can possibly be added to the 'multilingual matrix English/Kriol' rate. Nevertheless, none of the languages notably stand out in rate in the overall distribution of unilingual clauses. Another observation in Table 13 concerns the ML of the 632 multilingual clauses. Multilingual clauses with the Spanish ML are substantially more preferred than multilingual clauses with English or Kriol as the ML. Once again, ML Kriol and English are not combined in the multilingual clauses, it simply refers to either of the languages. Yet, keeping in mind the possibility that unilingual English/Kriol clauses may, in fact, be added to the number of multilingual clauses, with English/Kriol as an ML. Regardless of that, we can state that whenever CS occurs between an Anglophone language and Spanish, the ML is notoriously preferred in Spanish (87%).

Table 13.

Language distribution and proportion over the total amount of unilingual and multilingual clauses.

	amount	percentage
total_unilingual_clauses	**3214**	**100%**
total_unilingual_clause_english_or_kriol	1921	60%
total_unilingual_clause_spanish	1293	40%
total_multilingual_clauses	**632**	**100%**
total_multilingual_clause_matrix_english_or_kriol	83	13%
total_multilingual_clause_matrix_spanish	549	87%

8.3 CS attitudes and self-rated proficiency

We registered each participant's self-rated proficiency and CS attitudes in our Methodology Chapter. We can observe individual differences in English, Spanish, and

Kriol proficiency levels across all 18 participants. All proficiencies that were rated 'excellent' and the distribution among the participants are displayed in Table 14. The age of acquisition also differs, especially for Kriol. At the group level, we also see that CS attitudes are overall positive. Only in a few cases is the CS attitude negative. Despite this, all participants reported engaging at least 'sometimes' in CS and mainly reported that they engaged 'often' or 'always' in CS.

Table 14.

Self-rated proficiency excellency rates in the dataset (n=18).

Speaking & understanding proficiency 'excellent'	Amount of participants	Participants
All: English, Spanish, Kriol	4	glekan, leorey, keylem, adrtun
Two: English, Spanish	6	carcha, emaort, shecar, emmarc, jacbla, elamog
Two: English, Kriol	3	vivnov, jaiben, josnah
Two: Spanish, Kriol	0	-
One: English	4	zuruk, rongon, kiamai, daycas
One: Spanish	1	evedel
One: Kriol	0	-
Total:	**18**	

8.4 Linking sociolinguistic factors with grammatical outcomes

The chart following this section provides a comprehensive overview of all our participant's profiles and self-reported data, including proficiency, CS attitudes, and CS engagement (Figure 8). This chart also includes self-reported network data[23], where

---

[23] This network data was extracted from the language background questionnaire of these participants by Dr. Floor van den Berg, the postdoctoral researcher on the Crossing Language Borders project.

participants reported which languages they speak on a daily basis with people from their social network. Specifically: "Which of these languages do you use with this person in general?"

Several results can be observed when connecting the self-reported data with the production outcomes of the unilingual and multilingual clauses. When Kriol's proficiency is rated 4/4 ('excellent') in both *speaking* and *understanding*) (7 out of 18 participants), unilingual anglophone clauses are consistently produced more frequently than unilingual Spanish clauses, despite differences in the age of acquisition. In cases where unilingual anglophone clauses are produced more frequently than Spanish unilingual clauses, it is consistently observed that both Kriol and Spanish proficiencies are rated less than 'excellent' (in 4 instances). When both English and Spanish, and not Kriol, are rated 'excellent', Spanish unilingual clause production is always higher than English (anglophone) unilingual clause production (in 6 cases). These observations suggest a consistent relationship between proficiency and unilingual speech production.

Moreover, a hierarchy in the relationship between proficiency and production can be inferred. Noting that this is in the case of unilingual speech, and not in multilingual speech. Specifically, 'excellent' proficiency in Kriol (even in combination with 'excellent' proficiency in English and/or Spanish) corresponds to more unilingual anglophone outcomes. When English 'excellent' proficiency competes with solely Spanish 'excellent' proficiency, there are always more Spanish than anglophone unilingual outcomes. If English 'excellent' proficiency occurs with lesser proficiency in Spanish and Kriol, the unilingual outcomes will primarily be anglophone (reasonably). Notably, there is only one case (*evedel*) where English proficiency is not rated 'excellent'. In this case, only Spanish was rated 'excellent', yet the unilingual outcomes were overwhelmingly anglophone (98%). The prestigious status of Kriol may explain this hierarchical structure regarding proficiency-rated unilingual outcome rates. We will elaborate on this in our Discussion.

Although data on the participants' language attitudes were not collected, attitudes towards CS were. However, there seems to be no consistency between attitudes, self-reported CS engagement, and multilingual clause production.

In our dataset, we controlled for age, with one exception. It can be observed that the outlier in age (*elamog*, 76-year-old) produces more code-switches than anyone else

in the dataset. When comparing his percentage of multilingual clauses (52%) with, for instance, the participant with the second highest multilingual clause production (33%, 19-years-old *emaort*), we see a considerable difference. This suggests age is a potential influential factor on the proportional outcomes of unilingual and multilingual speech in this community.

Figure 8. *Individual speaker's profiles, including age, gender (M/F), language production outcomes, proficiency, age of acquisition, and social network.*

8.5 Preliminary network data observations

The questions regarding the participants' social network (egos) included elicitations on their CS behavior with their speaking partners of the network (the alters). The intensity of CS engagement between the ego and their alters is not shown in the chart above nor discussed in this study. However, we included ties indicating whether the alters also communicate with each other. This data is used in addition to our outcomes, as it became available during our study.

Preliminary observations suggest that participants use multiple languages in a single conversation with most alters, for instance indicated by the pink dot, representing Kriol, English, and Spanish. There is a considerable variation in language use both individually and across the dataset. Self-reported rates of CS engagement are likely relative for each speaker, based on their observations within their social network or beyond.

For instance, *adrtun,* an 18-year-old male speaker, as shown in the chart, produced 42 multilingual clauses, accounting for 24% of his total clause production (176 clauses). He reports 'sometimes' engaging in CS. In contrast, an 18-year-old female speaker, *glekan*, with the same proficiency rates, reports 'always' engaging in CS but has only a 6% rate of multilingual clause production. Both their social networks show that all alters know English, Spanish, and Kriol. In *adrtun*'s case, these alters also interact intensively with one another, whereas in glekan's case, these ties are not visible.

Similarly, we observe cases where the ego reports speaking one language (e.g., Spanish (red) or Kriol (yellow)) with their alter, while these alters interact with one another in at least one of these languages. For example, *leorey* speaks with four people in one language only (Spanish or Kriol), but these alters also interact. Interestingly, *leorey*, reports 'often' engaging in CS, and has a positive attitude towards it. This suggests that 'often' can be relative. The same applies to *daycas*, who reports 'always' engaging in CS and is positive about it, yet points out three people in her social network with whom she only conversates in Spanish.

Our data outcomes suggest that 'unilingual' conversations may occur, as in the case of *evedel*, who speaks all three languages, but chooses to speak mostly one

language with his conversational partner (*josnah*). While *josnah* reports using multiple Anglophone languages in his social network, his unilingual speech consists of 30% Spanish. This raises questions about whether he adapts to *evedel*, knowing that *evedel*'s first acquired language is Spanish.

When we compare *evedel*'s social network to anyone else's social network, we find that *elamog*, our outlier in age, has the exact same social network profile: four people with whom they speak with in Kriol, English, and Spanish, and three people with whom they speak English and Spanish, and no Kriol. Besides, they also match in Kriol and Spanish proficiency, and mostly English, and their ages of acquisition are fairly similar, with again a slight difference in English. Both males have positive CS attitudes and they both report to frequently engage in CS. In other words, their greatest difference lies in age. Remarkably, their speech production outcomes are rather the opposite: *evedel* rarely code-switched within a clause (intraclausal CS), or even in between unilingual clauses (interclausal CS). *Elamog*, on the other hand, code-switched roughly half of his speech at intraclausal level. His unilingual speech is about 65% versus 35% divided in Spanish and Anglophone, respectively. Generally considering his multilingual speech is much more dense than evedel's, or anyone else's speech, age is a factor that needs exploration in this community.

9. Discussion

In the present study, we aimed to identify proportions of unilingual versus multilingual speech in the spontaneous conversational speech of 18 speakers living in or near Orange Walk Town in Northern Belize. More specifically, we looked at the individual-level and community-level distribution of unilingual versus multilingual speech and the choice in ML (**Research Question 1**). Additionally, we sought to relate these linguistic outcomes to the participant's attitudes toward using multiple languages in a single conversation, their proficiency in each language, and we preliminarily explored their social network information (**Research Question 2**).

We fully transcribed and coded their video-recorded speech for our research. Each participant completed an extensive background questionnaire, from which we extracted their CS attitudes and other metalinguistic data. We provided the (socio)historical linguistic context of this community and reviewed previous literature on language attitudes in Northern Belize specifically. Furthermore, we applied the widely used MLF model, which explains and predicts the linguistic structures in which CS operates. It poses an asymmetry between the dominant ML, providing the morphosyntactic frame of the clause and the EL following this framework. By determining the ML of each clause, we aimed to identify whether there was a preference in ML choice at the individual or community level.

With this comprehensive background information, we aim to connect our study's results with previous findings in CS research and the sociolinguistic context of Northern Belize.

In this study, we identified 4010 clauses, of which 3214 (80%) were unilingual and 632 (16%) were multilingual. No marked preference for the choice in ML in unilingual clauses was determined. In contrast, in multilingual clauses, Spanish emerged as the preferred Matrix Language. Self-rated proficiency seems related to patterns in unilingual, but not multilingual clauses. The Spanish ML preference in multilingual speech may be related to positive attitudes toward Kriol, or the language use in social network of the speakers, however, the latter needs further exploration.

Research Question 1. What is the distribution of unilingual as opposed to multilingual speech in Northern Belize, and is there a preference for the Matrix Language (ML) in code-switched speech?

The proportions of unilingual and multilingual CS speech production are as follows. The total number of the 18 video-recorded participants is 5337 speech segments, of which 4279 (80%) are unilingual and 717 (13%) are multilingual, and the rest could be both ('flexible', 341 cases, 6%). Of this total number (5337), we identified 4010 clauses (75%), 1155 (22%) 'other' speech segments (without subject and verb). There were 172 instances (3%) that could be both ('flexible'), as some were unclear to the transcriber, e.g., due to inaudibility.

We now zoom in to the linguality of the 'clauses' (4010) and 'other' speech segments (1155). We found that 3214 out of total 4010 (80%) clauses were unilingual, meaning that the full clause was either Spanish or Anglophone (English, Kriol, or a mix of those two). In continuation, 632 out of total 4010 clauses (16%) were multilingual, and the rest (164 cases, 4%) was 'flexible' (unidentifiable). The percentages of the 'other' speech segments were differently divided: 1070 out of total 1155 speech segments (93%) were unilingual, and 85 out of total 1155 (7%) were multilingual. We could determine all 'other' speech segments on its linguality, so there were no 'flexible' instances.

We now concentrate solely on the unilingual (3214) and the multilingual (632) clauses produced in the dataset. Anglophone unilingual outcomes (1921 clauses, 60%), this includes undetected CS between English and Kriol, are slightly higher than Spanish (1291 clauses, 40%). In multilingual clauses, we found 539 instances (87%) in which the Matrix Language (ML) was Spanish, and solely 83 cases (13%) in which the ML was Anglophone.

Hypothetically, if CS occurred as much between English and Kriol as in Spanish-Anglophone clauses, then 632 out of 1921 unilingual Anglophone clauses were bilingual. If this were the case, then 1289 Anglophone unilingual clauses (49,9%) were produced, versus 1293 Spanish clauses (50,1%). Given that Anglophone unilingual clauses consist of two languages (say, those 1289 Anglophone clauses divided by two languages), we may argue then that unilingual Spanish is slightly preferred over Kriol and English. Either way, a clear language preference is absent in the unilingual outcomes.

In sum, we can observe a strong preference for Spanish being the ML in Anglophone-Spanish clauses. Despite the clear observation of Spanish ML preference in multilingual clauses, we generally see no considerable preference for the language used in unilingual clauses; patterns differ individually.

With Spanish being the *lingua franca*, we expected to find a more significant amount of unilingual Spanish clauses than those of Anglophone languages. Only seven of our 18 participants produced more unilingual Spanish than Anglophone clauses. One of these seven participants was a 76-year-old male, someone we consider an outlier of our dataset due to his age. We take into account that a significant part of the Anglophone clauses may be multilingual. Nonetheless, we found another five out of 18 cases where the proportions unilingual Anglophone versus Spanish were as follows: Anglophone 85% or higher, versus 15% or lower Spanish. Even when Anglophone clauses are partly multilingual in these cases, the number of unilingual Anglophone clauses still outweigh the number of Spanish unilingual clauses.

As for the multilingual outcomes, on the other hand, the observation for ML preference in multilingual speech was previously found by Carter et al. (2011). Namely, in the Miami corpus, in one of the three examined bilingual speech communities, about two-thirds of the multilingual clauses were with Spanish as an ML. The authors relate the ML preference for Spanish to proficiency, ethnic identity, and their participants' social network. Other observations on ML preference in multilingual clauses, such as in the Nicaraguan Creole English-Spanish data of Blokzijl et al. (2017), were explained by the relation between the ML and the language of prestige in the community. In our dataset, the participants were all, to a certain level, proficient in Kriol, English, and Spanish, with some individual differences. We will elaborate on these extralinguistic factors on the hand of our second research question.

Research Question 2. To what extent is there a relation between CS attitudes, proficiency, the languages used in participants' social networks, and the proportional outcomes of unilingual and multilingual speech?

When we relate these speech production outcomes with sociolinguistic factors such as CS attitudes, language proficiency and preliminary social network data, we first observe a correlation between unilingual data and language proficiency. No clear relation was found between participants' proficiencies and the multilingual outcomes. Participants who reported being less proficient in Spanish did not produce significantly fewer multilingual Anglophone-Spanish clauses than those with high Spanish proficiency. Self-rated proficiencies between Kriol, Spanish, and English varied individually. We observed a (presumably hierarchical) relationship between the language of the most unilingual clauses produced and proficiency levels. Namely, 'excellent' Kriol proficiency (even combined with 'excellent' English and/or Spanish proficiency) led to more unilingual Anglophone outcomes. When English 'excellent' proficiency *competed* solely with Spanish 'excellent' proficiency, more Spanish unilingual outcomes were observed. If English 'excellent' proficiency was paired with lower proficiency in Spanish and Kriol, the unilingual outcomes were (reasonably) primarily Anglophone. There was considerable individual diversity in unilingual outcomes; seven of eighteen participants had a higher rate of unilingual Spanish than unilingual Anglophone clauses. Although, some Anglophone unilingual clauses may be code-switched English-Kriol multilingual clauses.

We reviewed reports on language attitudes in Belize, which identify Kriol as the most prestigious language, often associated with Belizean identify, also in Northern Belize. Some of these studies were conducted in a language contact situation where Kriol and the minority language Garifuna or Mopan (Mayan) are spoken (Bonner, 2001; Ravindranath Abtahian, 2009, 2017; Gómez Menjívar & Salmon, 2018). In addition, these reports are based on speaking environments where Kriol is the *lingua franca*, and not Spanish, as in Northern Belize. This would suggest two competing elements between Spanish and Kriol in Northern Belize: (i) Kriol as a more prestigious language and (ii) Kriol as the *lingua franca* in the rest of Belize. Generalized language attitudes toward English remain unclear, as few studies have been conducted on this matter. Our participants, mostly young adults, grew up in a globalized environment, with more access to English-

rooted technology, such as social media, movies, podcasts, and so on, in comparison with older generations. Our auditory observations in this study suggest that English is frequently used in spoken speech, in contrast to previous reports, stating that English is mostly written (e.g., Holm, 1983; Balam, 2013). Reports on language attitudes toward the Spanish variety in Northern Belize are not overtly positive, but not so negative either (Balam, 2013; Bero, 2022).

However, these language attitudes are based on attitudes toward a single language, usually explored isolated from the context of the intense form of multilingualism that Belize offers. Our participants did not answer questions about their attitudes toward each of the languages individually, but we acquired their attitudes towards CS. Overall, most participants had a positive attitude toward CS (13 out of 18), some were neutral (2 out of 18) and some were negative (3 out of 18). There was no notable correlation between the CS attitudes in regard to the outcomes; those who had neutral or negative attitudes, did not engage less in CS (as observed in the speech outcomes) than those with a positive attitude. Besides this, our participants also reported on their frequency in CS engagement. Only three out of 18 reported to 'sometimes' engage in CS, the rest engages 'often' or 'always'. None of the speakers 'rarely' of 'never' code-switches. Notably, these three participants were not the same as the ones with negative attitudes. Relating these self-reported frequencies on CS, we observed no correlation in outcomes either; these three participants produced 52 (19%), 42 (24%), and 57 (33%) multilingual clauses (in which percentages are from the total amount of produced clauses). Given that we found that the average amount of the total produced multilingual clauses is 16%, these participants are, remarkably, above average.

Preliminary results on the social network, reported by the participants, show that the usage of single languages and language combinations strongly vary. Both in individual cases and across the dataset. Speakers always produced unilingual speech in both Spanish and Anglophone clauses with their conversational partners, never solely Spanish or Anglophone. We observed in the paired conversations, in which we know the proficiencies of each speaker, that in 8 out of 9 recordings, both speakers preferred the same language in unilingual speech. In some cases, this does not reflect the main

languages or language combinations used in their social network. This may suggest that these speakers naturally adapt their language use to the other speaker's preference, perhaps influenced by proficiency levels, language attitudes, or other factors.

Multiple factors may explain findings on language and CS attitudes in Northern Belize. First, Kriol's popularity is increasing in this region. Spanish is the *lingua franca* of Northern Belize, had been adopted by immigrants and passed on to the next generations over the past 150 years. It developed in relative isolation until the mid-20th Century when improved infrastructure spurred national and international migrations, creating a unique Northern Belizean Spanish (NBS) variety. Current speakers of this NBS sometimes perceive it as a 'broken' or 'lesser form', but attitudes are generally positive or neutral (Balam, 2013).

Since the 1970s and 1980s, attitudes toward Kriol slightly changed from overly negative (a 'broken' or 'lesser' form of English) to the status of a prestigious language, and the *lingua franca* of Belize (except for the Northern Districts) (Gómez Menjívar & Salmon, 2018). In the coastal areas of Belize, reports on Garifuna-Kriol speakers indicate Kriol as the dominant spoken language (e.g., Ravindranath, 2009, 2017). This contrasts with a few decades ago, when Garifuna was the most dominant spoken language in this community. Although Spanish in Northern Belize is not a minority language like Garifuna, a similar dominance shift in spoken language may be occurring. Census data from 2022 (Statistical Institute of Belize, 2022) shows a slight increase in Kriol speakers compared to 2010 (Census 2010, Statistical Institute of Belize, 2010), suggesting its growing prestige and role as an identity marker of being 'Belizean'.

Muysken (2013) distinguishes four types of CS, including backflagging, which involves discourse markers. Although Muysken (2013) relates this CS strategy mostly to heritage speakers, which our participants are not, we do understand that our participants use these as identity markers. In our analyses, we considered (other language) discourse markers peripheral to the otherwise unilingual segments. We found Anglophone discourse markers used within or around 52 Spanish unilingual speech segments. This was more than twice the amount of Spanish discourse markers within or around 21 Anglophone speech segments. This contrast highlights CS as a strategy for social marking (with ethnic or other social values), with Spanish and Kriol historically tied to mestizo and

Creole identities, respectively. In the naturalistic corpora of Fuller Medina (2021) of Northern and Western Belizean, she also reported on frequent production of Anglophone discourse markers peripheral to Spanish unilingual clauses (18% of her dataset). This is, after Anglophone noun insertions (57%) the second largest CS category in her corpus.

This social marking may explain why Spanish predominantly serves as the ML in this dataset. Rather than signifying Spanish's prestige or strong social, ethnic, or other value, this usage likely supports general language maintenance or continuation as the *lingua franca*. Indirectly, this is explained by Carter et al. (2011) in twofold; (i) the Spanish ML preference is a way to cope with countering strong (ethnic) identity marking of the uprise of prestigious Kriol, and (ii) the insertion of Anglophone elements into the clause serve as identity marking. In their follow-up study, Parafita Couto et al. (2014) examined the underlying regularities and differences to understand the uniformity in the ML choice in Carter et al.'s (2011) Welsh-English, and not in the Spanish-English corpus data. They analyzed community-level characteristics such as age of acquisition, proficiency and educational levels, national identity, and languages of the participants' social networks. These factors could determine if community characteristics explain community norms in multilingual speech. In the Welsh-English corpus, they found no correlation between age of acquisition or proficiency levels and ML preference but suggested it could be explained by identity marking, educational levels, and the languages used in the participants' social networks. From a grammatical perspective, it could also be argued that switching between VSO word order (Welsh) and SVO word order (English) is dispreferred. These explanations are not mutually exclusive; both could be contributing factors.

Limitations and future research

We recognize that the size of our dataset is limited, both in participants and the hours of recorded speech. Additionally, self-rated proficiencies and the age of acquisition of each language are somewhat divided and not consistently accurate or reliable. When speakers reported having learned the language from '3' years old, we assumed it to be 'from birth', as no lower age was pointed out in the entire dataset. However, we have no background on whether this was learned exactly from birth or perhaps slightly later. Besides this, self-rated proficiencies are always related to the participant's perception of their speaking

environment. For instance, a speaker might not rate himself 'excellent' in Spanish when a Central American Spanish variety is spoken in his social network, not the NBS variety. Alternatively, a speaker may identify their acrolectal form of Kriol as English, and not Kriol, and therefore rate their Kriol as 'poor' or 'fair', instead of 'good' or 'excellent', or vice versa, with English proficiency rating. For future research, it would be necessary to investigate these proficiencies quantitatively by conducting and developing language proficiency tests for Kriol, for instance. Another limitation is the fact that we do not have data on the individual language attitudes of our participants, so we were obliged to review recent studies on language attitudes toward Spanish, Kriol, and English. Although studies were conducted on this topic, which provided us insights in language attitudes of this Northern Belizean community in general, it would be better to have the information of our particular dataset. Also, the participants of this dataset were partly conscious of the aim of this study. The information sheet clearly stated what CS entails and why it is interesting to look at in Belize. It might be that CS occurred more often in our dataset than usual or that reports on CS attitudes in this study are more positive. Either way, the linguistic constructs in which the CS occurs are specific to this community, and therefore, we argue that the CS patterns remain the same. In some cases, the camera operator was (partly) present during the recording. For this reason, participants may have felt they could not speak freely at certain points, or may have behaved differently, as they were conscious of being observed ('the observer's paradox', Labov (1972)). One recording (*adrtun – elamog*) needed to be retaken several times, due to background noise and relocation. This may have influenced the spontaneity of their speech, although, again, we do believe that it would minimally influence CS patterns. Some limitations for corpus building need to be mentioned as well. First, all transcription and coding were done manually, a highly time-consuming task. We are aware that an error margin needs to be calculated. We followed our transcription conventions and kept track of the transcription issues meticulously. Nevertheless, each transcriber may perceive spoken speech differently, and during the hundreds of hours put into transcription and coding, inconsistencies may have entered the transcripts. During transcription, we also found instances where longer pauses in speech (or any other clear prosody-related boundaries) influenced our choices whether a speech segment was part of the clause, or whether we

count it as 'other' speech segment. Only in cases where there was a clear boundary, we separated those. While we only looked at the amount of clauses for our analysis, the proportions remain the same. However, it would be beneficial to establish conventions and take prosody into account for future research.

This study has provided several insights into proportions of CS within a community where CS is the norm, given the relatively high proportion of multilingual clauses (16%). It also shows how language proficiency, language and CS attitudes are related with grammatical outcomes. Nevertheless, our dataset is limited; therefore, our type of research yields further investigation on different fronts.

Besides enlarging the dataset, further exploring the individual language attitudes in this community in Northern Belize is of high priority. This includes examining the use of and attitudes toward the standard variety of English. Also, in comparison with Central American English and Standard American English. The use and language attitude of each language involved also needs further exploration, based on age and gender across different towns within this region.

For this study, we used an impressionistic auditory analysis to determine the difference between English and Kriol. Nevertheless, we could not define the difference between the languages on a clausal level. After consulting this challenge with Osmer Balam, who we cited more than once in this thesis, we concluded that a methodology to systematically distinguish Kriol from English on a clausal level is needed to conduct further research on CS between this variety of English and Kriol.

When observing our results, we are not in a position to question the current *lingua franca*. However, given that nowadays English is frequently used in spoken speech and no longer limited to written texts (also observed in our dataset), proportions of spoken English may require further exploration. Controlling for age possibly provides insights whether young Northern Belizeans tend to use more Anglophone speech than older speakers. If so, then unilingual and possibly multilingual speech patterns may vary between different age groups as well.

Future explorations in our dataset on intraclausal CS, particularly within grammatical conflict sites, such as determiner-noun-adjective constructions and bilingual compound verbs, can provide valuable insights into CS patterns within the community. Grammatical outcomes at these conflict sites may illuminate how individuals navigate multilingual speech. Additionally, results from Acceptability Judgment Tasks or other methods used to elicit comprehension, could complement and expand our understanding of CS patterns in naturalistic speech data. Lastly, from a cognitive perspective, investigating the cognitive mechanisms of CS may reveal whether multilinguals (specifically habitual code-switchers) differ in processing or inhibiting multilingual versus unilingual speech.

Suggestive further exploration

We conclude this study by suggesting further exploration of social inclusion in a community where CS is habitual and socially embedded. Our data shows that 13% of *all* produced naturalistic speech is multilingual, including clauses and 'other' speech segments. Moreover, none of the speakers in our dataset produced solely one language in their unilingual speech, meaning that interclausal and intersentential CS is common. Given that CS is standardized speech here, outsiders (e.g., national or international immigrants) who intend to blend into the community, will need to adapt to this standard. This final section offers 'food for thought' based on observations in the video-recordings and personal anecdotes from the study's speakers, without making assumptions or drawing definitive conclusions.

Our relatively small dataset suggests that the Northern Belizean speech community consists of individuals with variable language proficiencies, CS attitudes, and self-reported CS engagement. The use of individual languages and language combinations varies widely among the social networks among our participants. This indicates that the community comprises numerous smaller group speech profiles rather than it is a homogeneous speech community. One group may have an immigration background and be more surrounded by Spanish speakers. Another group may have learned Spanish at different stages of life or in different contexts, such as in church, school, among friends, or in other situations. The same applies to Kriol, which is not an

immigrant language and can only be learned in social situations (through speech) in Belize. Kriol is typically acquired later in life or in different linguistic contexts, as seen in our dataset. Another group may also be familiar with languages other than Spanish, Kriol, and English, influencing their language proportions, ML choice in unilingual or multilingual speech and CS frequency. These diverse profiles result in varied unilingual speech outcomes and follow, remarkably, similar outcomes in multilingual speech.

The sociolinguistic context and background of these multiple groups profiles explain these differences, and our study's outcomes represent solely the tip of the iceberg. Some speakers with self-rated low proficiency in Spanish still produced code-switched clauses with Spanish ML. This suggests that they not only acquire Spanish, but also learn to code-switch. Their aim appears to be participation in the community's social norm. Our dataset indicates a consensus on CS engagement, with participants considering CS normal, necessary, and even '*an art*'. Fuller Medina (2021:143) found a small percentage of false starts, filled pause, and other disfluencies in her naturalistic data from Northern and Western Belize. She argues that if the percentages were larger, then this would suggest a heavy social stigma mitigating CS. In our data, we found a number of instances of disfluency in multilingual speech: hesitations, corrections in the form of translations, and a considerable amount of fillers. Speakers with low Spanish proficiency produced ungrammatical constructions in Standard Spanish and likely also ungrammatical in NBS. They repeated other speaker's multilingual segments with hesitations. We also found instances of repetitive code-switched fragments or translations, possibly serving as sociolinguistic markers in multilingual speech. Instances of CS primed within the speaker (e.g., by discourse markers) or by their conversational partner were also noted.

This habitual social drive to code-switch warrants for further exploration of the speakers' social contexts. If encouraged to code-switch habitually, speakers might need to learn how to engage in CS or even acquire a community language through CS engagement. To what extent do these speakers adapt, in terms of language acquisition and even, in a certain way, CS acquisition, to participate in this community's practice? Will they follow the established CS patterns or undergo different stages of adaption? Moreover, what are the underlying social motivations for doing so, or why not?

CS acquisition involves (i) learning how to engage in CS while knowing multiple languages or (ii) acquiring a language through CS while simultaneously learning how to code-switch. This counts for both children and adults integrating in Northern Belizean society, such as national or international immigrants. Additionally, will established code-switchers (those being part of the community) adapt to accommodate CS learners? Future research could explore CS acquisition and the driving social motivations by examining CS engagement intensity in the community's social network and relating it to individual proficiencies. Investigating speech disruptions, disfluencies, social identity markers (e.g., discourse markers), grammatical inconsistencies (e.g., within-speaker variation in unilingual NBS grammar), and priming, could provide insights into how CS emerges and persists in communities, such as in Orange Walk Town in Northern Belize.

Bibliography

Al-Bataineh, H., & Abdelhady, S. (2020). Cree-English intrasentential code-switching: Testing the morphosyntactic constraints of the Matrix Language Frame model. *Open Linguistics*, *6*(1), 322–322. https://doi.org/10.1515/opli-2020-0017

Balam, O. (2013). Overt Language Attitudes and Linguistic Identities among Multilingual Speakers in Northern Belize. *Studies in Hispanic and Lusophone Linguistics*, *6*(2), 247–278. https://doi.org/10.1515/shll-2013-1150

Balam O. (2014). Notes on the history and morphosyntactic characteristics of Spanish in Northern Belize. *Kansas Working Papers in Linguistics*, 35, 79–94.

Balam (2016a). *Language use, language change and innovation in northern Belize contact Spanish*. ProQuest Dissertations & Theses.

Balam O. (2016b). Mixed verbs in contact Spanish: Patterns of use among emergent and dynamic bi/multilinguals. *Languages*, 1(3), 1–21.

Balam, O., & de Prada Pérez, A. (2017). Attitudes Toward Spanish and Code-Switching in Belize: Stigmatization and Innovation in the Spanish Classroom. *Journal of Language, Identity, and Education*, *16*(1), 17–31. https://doi.org/10.1080/15348458.2016.1260455

Balam, Osmer & Parafita Couto, M. Carmen (2019). Adjectives in Spanish/English Code-Switching: Avoidance of Grammatical Gender in Bi/Multilingual Speech. Spanish in Context, 16(2).

Balam, Osmer, Stadthagen-González, Hans, Rodríguez-González, Eva & Parafita Couto, M. Carmen (2022). On the grammaticality of passivization in bilingual compound verbs, *International Journal of Bilingualism,* 2023, 27:4, p. 415-431.

Baran, D. (2017). *Language in immigrant America*. Cambridge University Press.

Belize Kriol Project (1994). How fi Rite Bileez Kriol. Belize City: Belize Kriol Project.

Belize Kriol Project (1997). Bileez Kriol Glassary an Spellin Gide. Belize City: Belize Kriol Project.

Bero, T. J. (2022). *Spanish rhotic variation in Corozal Town, Belize.* Available from ProQuest Dissertations & Theses Global.

Bierings, Emma, Parafita Couto, M. Carmen & Mateo Pedro, Pedro (2019). Contrasting code-switching theories: Insights from Kaqchikel-Spanish code-switched nominal constructions. *UC Santa Cruz: Conference Proceedings*. Retrieved from https://escholarship.org/uc/item/6m86k80j

Blokzijl, Jeffrey, Deuchar, Margaret, & Parafita Couto, M. Carmen (2017). Determiner asymmetry in mixed nominal constructions: The role of grammatical factors in data from Miami and Nicaragua. *Languages*, 2(4).

de Boer, Meike M., & Heeren, Willemijn F. L. (2020). Cross-linguistic filled pause realization: The acoustics of uh and um in native Dutch and non-native English. *The Journal of the Acoustical Society of America*, *148*(6), 3612–3622. https://doi.org/10.1121/10.0002871

Boersma, P., & van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glot International*, 5(9/10), 341-347.

Bonner, Donna (2001). Garifuna children's language shame: Ethnic stereotypes, national affiliation and transnational immigration as factors in language choice in southern Belize. *Language in Society* 30(1). 81–96. doi: 10.1017/S004740450100104X

Brobbey, S. (2015). *Codeswitching on Ghanaian radio talk-shows: "Bilingualism as an asset."* The University of Bergen. Master thesis.

Bullock, Barbara E. & Toribio, A. Jacqueline. (eds.). 2009. *The handbook of linguistic code-switching*. Cambridge: Cambridge University Press.

Bulmer-Thomas, B., & Bulmer-Thomas, V. (2012). *The economic history of Belize : from the 17th century to post-independence* (first edition.). Cubola Books.

Cal, Angel and Fuller Medina, N. (2017). Revitalization of Yucatec Maya in Northern Belize: Training Yucatec Maya teachers to teach in primary schools of Maya communities. Unpublished ms. University of Belize.

Cardona Ramirez, M.A. (2010). La Fonética del Español en Belice. In (ed.) M.A. Quesada Pacheco, Lingüística Iberoamericana El español hablado en América Central: Nivel Fonético. 39, 20-54.

Carter, D. M., Broersma, M., Donnelly, K., & Konopka, A. E. (2018). Presenting the Bangor Autoglosser and the Bangor Automated Clause-Splitter. *Digital Scholarship in the Humanities*, 33(1), 21-28. https://doi.org/10.1093/llc/fqw065

Carter, Diana, Davies, Peredur, Deuchar, Margaret, & Parafita Couto, M. Carmen (2011). A Systematic Comparison of Factors Affecting the Choice of Matrix Language in Three Bilingual Communities. *Journal of Language Contact*, *4*(2), 153-183.

Crible, L., Degand, L., & Gilquin, G. (2017). The clustering of discourse markers and filled pauses: A corpus-based French-English study of (dis)fluency. *Languages in Contrast*, *17*(1), 69–95. https://doi.org/10.1075/lic.17.1.04cri

Comrie, B., Haspelmath, M. & Bickel, B. (2015). The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses by www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf.

Daugaard-Hansen, F. (2009). *'Coming home': The return and reintegration of Belizean returnees from the United States to Belize, Central America* (Order No. 3480443). Available from ProQuest Dissertations & Theses Global. (899274404).

Decker, Ken (2013). *The Song of Kriol: a Grammar of the Kriol language of Belize*. Electronic edition 2013. Belize Kriol Project 2005. SIL International: Dallas, TX.

Deuchar, M. (2006). Welsh-English code-switching and the Matrix Language Frame model. *Lingua*, 116(11), 1986-2011. https://doi.org/10.1016/j.lingua.2005.03.008

Deuchar, M. (2012). Code-switching. In Chapelle, Carol A. (ed.), The Encyclopedia of Applied Linguistics, 657-664. Wiley Blackwell.

Deuchar, M., Davies, P., Herring, J. R., Parafita Couto, M. C., Carter, D (2014). Building Bilingual Corpora. In Thomas, E. M., & Mennen, I. (Eds.), *Advances in the Study of Bilingualism* (pp. 93–110). Multilingual Matters. https://doi.org/10.21832/9781783091713-008

Deuchar, M., Donnelly, K., & Webb-Davies, P. (2018). Building and using the Siarad Corpus: Bilingual conversations in Welsh and English. *Studies in Corpus Linguistics* (No. 81). John Benjamins Publishing Company. https://doi.org/10.1075/scl.81

Dewaele, J.-M., & Li, W. (2014). Attitudes towards code-switching among adult mono- and multilingual language users. *Journal of Multilingual and Multicultural Development*, 35(3), 235-251. https://doi.org/10.1080/01434632.2013.859687

Edwards, J. A., & Lampert, M. D. (1993). *Talking data: transcription and coding in discourse research*. Lawrence Erlbaum Associates.

Escure, Geneviève (1997). *Creole and Dialect Continua: Standard acquisition processes in Belize and China (PRC)* (1st ed.). John Benjamins Publishing Company.

Escure, G. (2004). Belize and other central American varieties: morphology and syntax. In B. Kortmann & E. Schneider (Ed.), *A Handbook of Varieties of English: A Multimedia Reference Tool. Volume 1: Phonology. Volume 2: Morphology and Syntax* (pp. 1709-1736). Berlin, Boston: De Gruyter Mouton. https://doiorg.ezproxy.leidenuniv.nl/10.1515/9783110197181-106

Escure, G. (2006). Belize: Language situation. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics*, 2nd ed., Vol. 1, pp. 722-723. Elsevier Ltd. https://doi.org/10.1016/B0-08-044854-2/01772-7

Escure, G. (2013). Belizean Creole. In S. M. Michaelis, P. Maurer, M. Haspelmath, & M. Huber (Eds.), *The Survey of Pidgin and Creole Languages* (Vol. 1, pp. 92-100). Oxford: Oxford University Press.

Escure, G., & Schwegler, A. (2004). Garifuna in Belize and Honduras. In G. Escure & A. Schwegler (Eds.), *Creoles, Contact, and Language Change: Linguistic and Social Implications* (pp. 35-65). Amsterdam: John Benjamins Publishing Company.

Eppler, Duran E., Luescher, Adrian & Deuchar, Margaret (2017). Evaluating the predictions of three syntactic frameworks for mixed determiner-noun constructions. *Corpus Linguistics and Linguistic Theory*, 13(1), 27-63.

Fairchild, Sarah & Van Hell, Janet. 2015. Determiner-noun code-switching in Spanish heritage speakers. Bilingualism: Language and Cognition. Advance online publication. DOI: 10.1017/SS1366728915000619.

Fant, C. G. M., & Tatham, M. A. A. (1975). *Auditory analysis and perception of speech*. London: Academic Press.

Forker, D. (2019). Sanzhi–Russian code switching and the Matrix Language Frame model. *The International Journal of Bilingualism : Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, *23*(6), 1448–1468. https://doi.org/10.1177/1367006918798971

Fuller Medina, N. (2016). *Language mixing in northern and western Belize: A variationist comparative approach*. Doctoral dissertation. University of Ottawa, Canada.

Fuller Medina, N. (2020a). How bilingual verbs are built: evidence from Belizean varieties of contact Spanish. *Canadian Journal of Linguistics*, 65(1), p. 122-132. DOI: https://doi.org/10.1017/cnj.2019.1

Fuller Medina, N. (2020b). Belizean varieties of Spanish: Language contact and plurilingual practices. *Anuario de Estudios Centroamericanos, Dossier on Belize*, Vol. 46., p. 1-29. DOI: 10.15517/aeca.v46i0.42202

Fuller Medina, N. (2021). Lo que hacen mix es el Kriol y el English: How Spanish speakers reconcile linguistic encounters with English and Kriol in Belize. In G. Leung & M. Loschky (Eds.), *When Creole and Spanish Collide: Language and Cultural Contact in the Caribbean* (pp. 126-154). Brill Caribbean Series. Brill.

Gardner-Chloros, P. (2009). Sociolinguistic factors in code-switching. In B. Bullock & A.J. Toribio (Eds.), The Cambridge handbook of linguistic code-switching (pp. 97–113). Cambridge University Press.

Gómez Menjívar, J. C. & Salmon, W. N. (2018). *Tropical Tongues Language Ideologies, Endangerment, and Minority Languages in Belize*. The University of North Carolina Press.

Greene, L. A. (1999). *A grammar of Belizean creole : compilations from two existing United States dialects*. Peter Lang.

Gullberg, Marianne, Indefrey, Peter & Muysken, Pieter (2009). Research techniques for the study of code-switching. In: B.E. Bullock & A.J. Toribio (eds.), The Cambridge Handbook of Linguistic Code-switching. Part I, chapter 2, p. 21-39.

Holm, John (1977). 'Miskito words in Belize Creole'. Belizean Studies 5(6):1-19.

Holm, John (1978). The Creole English of Nicaragua's Miskito Coast: Its Sociolinguistic History and a Comparative Study of its Lexicon and Syntax. PhD thesis, University College, University of London. Ann Arbor: University Microfilms International.

Holm, J. (1983). *Central American English: Vol. v.T2* (1st ed.). John Benjamins Publishing Company.

Huddlestone, K., & Nel, J. (2012). Analysing Afrikaans-English bilingual children's conversational code switching. *Stellenbosch Papers in Linguistics*, *41*(1), 29–53. https://doi.org/10.5774/41-0-132

Ihemere, K. U. (2016). In support of the Matrix Language Frame Model: evidence from Igbo-English intrasentential code-switching. *Language Matters: studies in the languages of Africa*, vol. 47 (1), p.105-107.

Jones, D. A. (2023). A review and critical analysis of qualitative methodologies and data-collection techniques used for code-switching research. *American Journal of Qualitative Research*, 7(2), 53-72. https://doi.org/10.29333/ajqr/12988

De Jong, N. H., Pacilly, J. J. A., & Heeren, W. F. L. (2021). PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically. *Assessment in Education: Principles, Policy & Practice*, 28(4), 456-476. https://doi.org/10.1080/0969594X.2021.1951162

Joshi, Aravind (1985). Processing of sentences with intrasentential code switching. In D.R. Dowty, L. Kattunen, & A.M. Zwicky (eds), Natural language parsing: psychological, computational and theoretical perspectives (pp. 190–205). Cambridge: Cambridge University Press.

Khan, A. A., Khalid, A., & Saleem, T. (2023). The evidence of Embedded Language islands: the case of Pashto-English codeswitching. *International Journal of Multilingualism*, *ahead-of-print*(ahead-of-print), 1–24. https://doi.org/10.1080/14790718.2023.2212907

Kniaź, M., & Zawrotna, M. (2021). Embedded English verbs in Arabic-English code-switching in Egypt. *The International Journal of Bilingualism : Cross-Disciplinary, Cross-Linguistic Studies of Language Behavior*, *25*(3), 622–639. https://doi.org/10.1177/1367006920976909

Koenig, Edna L. (1975). Ethnicity and language in Corozal District, Belize: An analysis of code-switching. Doctoral dissertation, University of Texas, Austin.

Królikowska, M., Bierings, E., Beatty-Martínez, A., Navarro Torres, C, Dussias, P & Parafita Couto, M. C.(2019). Gender assignment strategies within the bilingual determiner phrase: Four Spanish-English communities examined. Poster presented at the 3rd conference on Bilingualism in the Hispanic and Lusophone World. Leiden: The Netherlands.

Labov, William (1972). Sociolinguistic patterns. Philadelphia: University of Pennsylvania Press.

Le Page, Robert B. (1992). You can never tell where a word comes from: Language contacts in a diffuse setting. In Ernst Håkon Jahr (ed.), Language contact, theoretical and empirical studies, 71-101. Berlin: Mouton de Gruyter.

Le Page, Robert B. & Tabouret-Keller, Andrée (1985). *Acts of identity: Creole-based approaches to language and ethnicity*. New York: Cambridge University Press.

Macdonald, S. (2024). In defence of Llanito: Gibraltar in a state of linguistic transition. *The Round Table*, *113*(3), 289–294. https://doi.org/10.1080/00358533.2024.2363044

McWhorter, J. H. (2018). *The creole debate*. Cambridge University Press.

Matras, Y. (2009). *Language contact*. Cambridge University Press.

Mufwene, S. S. (2001). *The ecology of language evolution*. Cambridge University Press.

Munarriz, Amaia. & Parafita Couto, M.Carmen. 2014. ¿Cómo estudiar el cambio de código? Incorporación de diferentes metodologías en el caso de varias comunidades bilingües, Lapurdum. Basque Studies Review XVIII: 43-73.

Muysken, P. C. (2000). *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: Cambridge University Press.

Muysken, P. C. (2013). Language contact outcomes as the result of bilingual optimization strategies, *Bilingualism: Language and Cognition*, 16(4), p. 709–730.

Myers-Scotton, C. (1997). *Duelling languages: Grammatical structure in codeswitching*. Oxford: Oxford University Press.

Myers-Scotton, C. (2002). *Contact Linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

Nguyen, T. (2015). *Code Switching: A Sociolinguistic Perspective*. Hamburg: Anchor Academic Publishing.

van Osch, Brechje, Parafita Couto, M. Carmen, Boers, Ivo, & Sterken, Bo. (2023). Adjective position in the code-switched speech of Spanish and Papiamento heritage speakers in the Netherlands: Individual differences and methodological considerations. *Frontiers in Psychology*, *14*, 1136023–1136023. https://doi.org/10.3389/fpsyg.2023.1136023

Parafita Couto, M. Carmen (2005). *Focus at the Interface*. ProQuest Dissertations & Theses.

Parafita Couto, M. C., Davies, P., Carter, D., Deuchar, M. (2014). Factors Influencing Code-Switching. In Thomas, E. M., & Mennen, I. (Eds.) *Advances in the Study of Bilingualism* (pp. 111–138). Multilingual Matters. https://doi.org/10.21832/9781783091713-009

Parafita Couto, M. Carmen, Bellamy, K., & Ameka, F. K. (2023). Theoretical Linguistic Approaches to Multilingual Code-Switching. In J. Cabrelli, A. Chaouch-Orozco, J. González Alonso, S. M. Pereira Soares, E. Puig-Mayenco, & J. Rothman (Eds.), *The Cambridge Handbook of Third Language Acquisition* (pp. 403–436). chapter, Cambridge: Cambridge University Press.

Pérez, D., & Sippola, E. (Eds.). (2021). *Postcolonial language varieties in the Americas*. De Gruyter. https://doi.org/10.1515/9783110723977

Poplack, Shana (1980). Sometimes I'll start a sentence in Spanish y termino en Español: Toward a typology of code-switching. *Linguistics*, 18, 581-618.

Poplack, Shana & Meechan, Marjory (1998). Introduction: How Languages Fit Together in Codemixing. *International Journal of Bilingualism* 2(2), 127-138.

Rahimi, M. & Dabaghi, A (20130. Persian-English code-switching: A test of the Matrix Language Frame (MLF) model. System, 41, 322-351.

Ravindranath Abtahian, M. (2009). Language shift and the speech community: Sociolinguistic change in a Garifuna community in Belize. Doctoral dissertation, University of Pennsylvania.

Ravindranath Abtahian, M. (2017). Language shift, endangerment and prestige Kriol and Garifuna in Hopkins, Belize. *Journal of Pidgin and Creole Languages*, *32*(2), 339–364. https://doi.org/10.1075/jpcl.32.2.05rav

Rushton, E. A. C. (2014). *"under the shade i flourish": an environmental history of northern belize over the last three thousand five hundred years*. ProQuest Dissertations Publishing.

Schneider, B. (2021). Creole prestige beyond modernism and methodological nationalism: Multiplex patterns, simultaneity and non-closure in the sociolinguistic economy of a Belizean village. *Journal of Pidgin and Creole Languages*, *36*(1), 12–45. https://doi.org/10.1075/jpcl.00068.sch

Shoman, Assad. (2010). Reflections on ethnicity and nation in Belize. Documento de trabajo 9, 1-61.

Statistical Institute of Belize (2010). *Census of Belize 2010*.

Statistical Institute of Belize (2022). *Census of Belize 2022*.

Straughan, J. F. (2004). *Belizean immigrants in Los Angeles*. ProQuest Dissertations & Theses.

The World Bank, World Development Indicators (2024). Retrieved from Population, total - Belize | Data (worldbank.org).

Torres Cacoullos, R., & Vélez Avilés, J. (2023). Mixing adjectives: A variable equivalence hypothesis for bilingual word order conflicts. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.22038.tor

Vanhaverbeke, M., & Enghels, R. (2021). Diminutive constructions in bilingual speech: A case study of Spanish-English codeswitching. *Belgian Journal of Linguistics*, *35*(1), 183-. https://doi.org/10.1075/bjl.00068.van

Velupillai, V. (2015). *Pidgins, Creoles and Mixed Languages: An Introduction* (1st ed., Vol. 48). John Benjamins Publishing Company.

Young, Colville N. (1995). *Language and education in Belize* (2nd ed.). National Printers.

APPENDIX A - INFORMATION SHEET

Universiteit Leiden
The Netherlands

Information Sheet

CROSSING LANGUAGE BORDERS: A QUEST FOR THE HUMAN LANGUAGE CAPACITY IN WEST AFRICA AND CENTRAL AMERICA

[Starting date of the project]

Dear participant,

First of all, thank you for participating. You will be taking part in the **Crossing language borders: A quest for the human language capacity in West Africa and Central America** research project conducted by Renzo Alberto Ego Aguirre Santa Cruz, PHD student, under supervision of M. C. Parafita Couto and Felix Ameka at Leiden University. Before the research project can begin, it is important that you read the information regarding the procedures we will conduct. Please, make sure to read this document carefully.

About the project
This project will focus on situations in which two or more people mix two or more languages when they speak: a phenomenon called code-switching or code-mixing. This phenomenon is something common around the world, but most of the time it is unfortunately discouraged. Nevertheless, as for Belize —and West Africa— , this is not the case. That is the reason why it is so important to understand the linguistic, social and other factors surrounding this interesting phenomenon that occurs in these communities like yours. You, as a participant, will help us to have a better understanding in the situation in Belize, but our colleagues from the University of Amsterdam will also work in Benin for having a broader idea of this situation.

About what you are required to do
Before recording begins, we will ask you to answer some questions related to your age, gender, occupation and neighborhood. We will also ask more detailed questions about languages you speak, your exposure to these languages, contexts in which you use these languages, etc. We will observe you interacting with people and record your spontaneous speeches. We will also ask you some questions related to social network. Then your conversations with one or two other participants will be recorded. The conversations will take around 45 minutes. Moreover, all the participants of this research will be compensated with 20BZD.

About potential adverse effects
The risks of participating in this research are no greater than in everyday situations. Previous experience in similar research has shown that no or hardly any discomfort is to be expected for participants. Please remember that you will participate in this research project on a voluntary basis. This means you are totally free to stop taking part of it at any stage. This will not have any consequences and you will not be obliged to finish the procedures described above. You can always decide to withdraw your consent later on. If you decide to stop, you can send an email with a short message indicating that you want your personal data to be removed. It is also important to mention that ethical approval has been acquired from the Leiden University Faculty of Humanities Ethics Committee for this research.

About treatment of the data
The information gathered over the course of this research will be used for further analysis and publication in scientific journals only. Your personal details will not be used in these publications. Video and audio recordings will also never be shown in public without your explicit consent. You can provide such consent on the informed consent form if you wish to do so. The research data will be stored for a period of 10 years. The personal data will only be stored as long as it is necessary for the research and will be deleted as soon as possible, unless you explicitly provide consent to store it for a longer period. You can provide such consent on the informed consent form if you wish to do so.

For further information
For further information on the research, now or in the future, you can turn to Renzo Alberto Ego Aguirre Santa Cruz (telephone number: +501-322-2280, e-mail: r.a.ego.aguirre.santa.cruz@hum.leidenuniv.nl; address: 14 Main Street, Orange Walk Town, Belize C.A.). Besides, if you have any complaints on this research, you can turn to the secretary of the Ethics Committee of the Faculty of Humanities of Leiden University, Myrte Vos, e-mail: ethics@hum.leidenuniv.nl, phone number: +31 71 527 2092

APPENDIX B – CONSENT FORM

Universiteit Leiden
The Netherlands

Informed Consent Form

'I hereby declare to have been informed in a way that was understandable to me, on the nature and method of the research, as was also laid out in the Information Sheet '**Crossing Language Borders: A Quest For The Human Language Capacity In West Africa And Central America**'. My questions have been answered satisfactory.

I realize that participation in this research is on an entirely voluntary basis. I retain the right to revoke this consent without having to provide any reasons for my decision. I am aware that I am entitled to discontinue the research at any time, and that I can always withdraw my consent after the research has ended. If I decide to stop or withdraw my consent, all the information about me gathered up until then will be permanently deleted.

The information gathered over the course of this research will be used for further analysis and publication in scientific journals only. Your personal details will not be used in these publications. Video and audio recordings will also never be shown in public without your explicit consent. The research data will be stored for a period of 10 years. The personal data will only be stored as long as it is necessary for the research and will be deleted as soon as possible, unless you explicitly provide consent to store it for a longer period.

If I would like to receive further information on the research, now or in the future, I can turn to Renzo Alberto Ego Aguirre Santa Cruz (telephone number: +501-322-2280, e-mail: r.a.ego.aguirre.santa.cruz@hum.leidenuniv.nl; address: 14 Main Street, Orange Walk Town, Belize).

If I have any complaints on this research, I can turn to the secretary of the Ethics Committee of the Faculty of Humanities of Leiden University, Myrte Vos, e-mail: ethics@hum.leidenuniv.nl, phone number: +31 71 527 2092.

I consent to:
- participate in this research   yes / no
- video/audio recordings being made   yes / no
-video/audio recordings being used for analyses   yes / no
-video/audio recordings being shown in public (e.g. conferences) if necessary   yes / no

'I have provides information on the research. I hereby declare myself willing to, now or in the future, answer any questions on the research to the best of my ability. '

Signed:

....

Name participant   Signature   Name researcher   Signature

....

Date

APPENDIX C – TRANSCRIPTION CONVENTIONS

Issue	Date	Solved yes / no	Solution / convention	Date	Annotation in transcription
Here we can write coding conventions / issues that we find and need to be solved or discussed with the researchers team / etc.					
When do we split the utterance, when is it a "pause"?	30-03-2024	yes	We divide the 'speech chunks' into clauses (for starters: every time we find a new verb phrase, we select it separately). Example: | "Every time we want to transcribe a clause," | "we split it into smaller pieces that" | "are smaller than a full utterance" | (|= separate selection in ELAN). Even if the pauses are large in between words of the same clause, they would still fall under the same clause.	10-4-2024	N/A
How to / do we annotate laughs?	30-03-24	yes	When it is part of a clause (it interrups it, or it starts/ends with it), you note it down as *laughs*. Otherwise you do not mark it.	10-4-2024	*laughs*
How to / do we annotate coughs / clearing throat?	30-03-24	yes	When it is part of a clause (it interrups it, or it starts/ends with it), you note it down as *coughs* or *clears throat*. Otherwise you do not mark it.	10-4-2024	*coughs* / *clears throat*
How to annotate hesitation (pause in between the same word)?	30-03-24	yes	Mark the hesitation with a hypen (-)	10-4-2024	h- he- hesita- hesitation
How to annotate repetition?	30-03-24	yes	Write down the repetition as they say it	10-4-2024	write write write it down
How to annotate small natural pause in between words (….milliseconds - ….ms)?	30-03-24	yes	We do not annotate this. We will look at individual cases when it seems relevant for the study.	10-4-2024	N/A
How to annotate longer pause in between words (… ms -… ms)?	30-03-24	yes	We do not annotate this. We will look at individual cases when it seems relevant for the study.	10-4-2024	N/A
How to annotate words that are unrecognizable?	30-03-24	yes	Write down UNCLEAR in your transcription and write it down in the separate Excel file 'Spreadsheet Belize transcription issues' under the time it occurs. Write down in the comments if something is unusual.	10-4-2024	Create a comment in the Spreadsheet Belize Transcription Issues file
How to annotate an utterance that is unrecognizable?	30-03-24	yes	Write down UNCLEAR in your transcription and write it down in the separate Excel file 'Spreadsheet Belize transcription issues' under the time it occurs. Write down in the comments if something is unusual.	10-4-2024	Create a comment in the Spreadsheet Belize Transcription Issues file
How to annotate fillers? (they are language specific)	30-03-24	yes	Write those down as they say it	10-4-2024	eh/ehm/uh/uhm
How to annotate (or later code) backchannels?	30-03-24	yes	Write those down as they say it	10-4-2024	hmm mm/uhu/ hmm
How to annotate half a word or when the clause gets disrupted?	30-03-24	yes	Mark the word-break with a hyphen (-). The same counts for a disrupted clause (by laughing/coughing/the speaking partner's speech etc.)	10-4-2024	half a wo-, When the clause gets disrupted by a-

Figure 8. *Transcription conventions for this study.*

Transcriber	File name & duration	Time	Issue	Comment	Second transcriber	Time	Comment
Emma Bierings	adrtunelamog171223ow	00:00:04	UNCLEAR	no suggestion	Simon Claassen	00:00:04	"okay, sigue ya"
	19min38s	00:00:11	UNCLEAR	no suggestion		00:00:11	"útlima vez que había mi mamá fue
		00:00:18	Spanish/ English: Octuber	How to annotate?		00:00:19	"de- when she came last time que-
		00:00:24	UNCLEAR	Spanish		00:00:24	"...la última vez de que I saw her w
		00:00:34	Spanish/ English: Chrustening	(christening)		00:00:31	"para su bauptismzo de ehm de la r
		00:00:48	*laughs*	How to annotate?		00:00:34	"for christening"
		00:01:00	"your pierents"	(your parents)		00:00:37	"feliz de que haba ella va a vernir a
		00:01:36	*laughs*	How to annotate?		00:01:00	"es algo que tus padrientes vienen,
		00:02:20	yeah la familia cuan- they come toget	"cuando" merged with "they" (Spanish-English)		00:01:04	"y entonces es una alegríiazo eso p
		00:02:47	UNCLEAR	name of school?		00:02:47	"uhm quiero conseguirme el degre
		00:03:13	con tu "esperience"	"experience" Spanish/ English		00:03:21	"weekend okay"
		00:03:24	yo quiero hacer "knuw"	"know" Spanish/ English		00:03:24	"yo quiero hacer knuow"
		00:03:39	UNCLEAR	"pues yeah there we don't have my papers in there so instead, instead for wait pero"		00:03:39	"pues yeah they already have my p
		00:04:04	más happiest	Spanish/ English how to annotate?		00:03:48	"ah quieres que hablanr un poco m
		00:04:31	UNCLEAR	name of a person		00:04:17	"no sólo tu madre mother viene"
		00:07:24	UNCLEAR	Speaker 2 is unclear (one or few words)		00:04:31	"Julita"
		00:07:30	UNCLEAR	"cachito"?		00:05:47	"y este ehm sé que hay unas cosas
		00:07:48	UNCLEAR	"when we busca a mi mamá"?		00:05:59	"okay, ellos quieren hacer tickets d
		00:08:40	UNCLEAR	person's name tío (few times)		00:06:04	"okay, entonces ellos que hacen ti
		00:09:54	UNCLEAR	"from finish she always puts.." ?		00:07:24	"yeah but"
		00:12:05	UNCLEAR	village name?		00:07:30	"no cada ratito pueden venir"
		00:12:20	* clicks with tongue*	How to annotate?		00:07:37	"a Chetoumal"
		00:12:58	P1 and P2 both talking UNCLEAR	no suggestion		00:07:48	"...'cause we're gonna go find ehm
		00:13:36	UNCLEAR	sportsterm?		00:07:59	"tú vas a ir a, you're going to Chete
		00:18:01	UNCLEAR	"on the list of permanancy"? (one word)		00:08:02	"tú vas a ir a Chetoumal"
		00:18:31	UNCLEAR	"qué va hacer happen?"		00:08:40	"tío Aléin"

Figure 9. *Transcription issues log file, in which each transcriber noted down any unusual or unrecognizable speech part (see link in Appendix E).*

Spreadsheet TRANSCRIPTION PER RECORDING BELIZE DATA

File name	Transcribed by	Start date	File length	Minutes transcribed	Finished yes/pending/no	Comments	1st Checked yes / no	Checked by	Date
keylemleorey121223ow	Emma (ETJB)	14-03-24	0:14:50	14min50s	yes	73 transcription issues	yes	Simon (SAC)	22-04-24
adrtunelamog171223ow	Emma (ETJB)	31-03-24	0:19:38	19min38s	yes	29 transcription issues	yes	Simon (SAC)	5-07-24
rongonzuruk241123ow	Emma (ETJB)	2-04-24	0:15:24	15min24s	yes	42 transcription issues	pending	Simon (SAC)	
carchaemaort171223ow	Emma (ETJB)	5-04-24	0:19:01	19min01s	yes	50 transcription issues	no		
daycasjaiben221123ow	Emma (ETJB)	14-04-24	0:15:46	15min46s	yes	88 transcription issues	yes	Simon (SAC)	7-06-24
jacblakiamai211123ow	Emma (ETJB)	20-04-24	0:15:14	7min15s	pending	28 issues so far			
emmarcshecar181223np	Renzo	27-04-24	0:23:47	17min04s	pending	179 issues so far			
glekanvivnov291123ow	Kiki	25-03-24	0:16:33	16min33s	yes	107 transcription issues	no		
evedeljosnah161123ow	Kiki	24-04-24	0:13:05	13min05s	yes	99 transcription issues	no		

Figure 10. *Transcription progress file*

APPENDIX E – ADDITIONAL LINKS

Link to background questionnaire

Background Questionnaire Belize

Link to ELAN Coding Guide

Elan Coding Guide & Manuals

Link to ELAN transcripts

Transcriptions used for thesis

Link to Excel coded transcripts, analyses, and metalinguistic data

version 2 Coding sheet ETJB ResMA thesis.xlsx

Link to transcription issues file

Spreadsheet Transcription issues Belize project.xlsx

Link to Crossing Language Borders project

Crossing Language Borders - ACLC - University of Amsterdam (uva.nl)