Available online at www.sciencedirect.com

ScienceDirect

Lingua 327 (2025) 104037

Lingua

www.elsevier.com/locate/lingua

An exploratory analysis of code-switching and borrowing in a corpus of Zimbabwean English

Faith Chiedza Chapwanya *, Joanine Hester Nel

Department of Afrikaans, University of Pretoria, South Africa

Received 4 March 2025; revised 1 September 2025; accepted in revised form 1 September 2025; available online 30 September 2025

Abstract

This paper explores patterns of English-Shona code-switching and borrowing in a corpus of Zimbabwean English to determine the types, frequencies, and functions of these features. Results indicate that intrasentential code-switching was the most frequent type of switching followed by intersentential code-switching and then lexical borrowing. The least frequent type of switching was loanword adaptations. The analysis revealed that spoken registers had the highest frequency of switching compared to written registers. The informal and spontaneous nature of spoken registers were cited as motivations for the high frequency. Subsequently, multi-word switches occurring in the same place were attested more, which can be linked to their higher frequency of occurrence in public dialogues. Social functions, namely accommodation and solidarity, had the highest frequencies whilst switching at a time of vagueness or uncertainty was the second highest function of code-switching. Other functions of code-switching attested in the Zimbabwean English corpus, in the order of frequency, include repetition or reiteration, asking questions, identity marking, code-switching to show emotions, code-switching as discourse marking, clarification, and emphasis. Analyzing the functions of code-switching in the Zimbabwean English corpus and determining the frequency of each function is crucial for understanding the linguistic-, cognitive-, and sociocultural dimensions of bilingualism in Zimbabwe.

Keywords: Code-switching; Zimbabwean English; Shona; Borrowing; Lexical borrowing; Loanword adaptations; Frequencies and functions

1. INTRODUCTION

One of the effects of globalization and migration is the mixing of people from different cultures and language backgrounds, resulting in language contact (Myers-Scotton, 1993a). Code-switching (CS), one of the by-products of language contact, has been studied extensively and globally, using different language pairs and from a variety of approaches (e.g. Blom and Gumperz, 1972; Gumperz, 1982; Poplack, 1980; Wardhaugh, 1992; Myers-Scotton, 1993a, 1993b). For the Global South, which includes Zimbabwe, language contact is a reality, leading to CS, borrowing,

* Corresponding author at: Department of Afrikaans, University of Pretoria, Lynnwood Road, Hatfield, Pretoria 0002, South Africa.
E-mail addresses: u12167160@tuks.co.za (F.C. Chapwanya), joanine.nel@up.ac.za (J. Hester Nel).

https://doi.org/10.1016/j.lingua.2025.104037
0016-7037/

and codemixing. CS and borrowing are of primary interest to this study as a result of the fact that bilingualism and multilingualism are the norm in many African countries. This study explores CS and borrowing patterns in a corpus of Zimbabwean English (ZimE). The term "Zimbabwean English" is used to refer to the English that is used by Shona first language (L1) speakers of English as a second language (L2). The study aligns with the pluricentric approach to World Englishes which acknowledges that the different contexts and environments in which English is used worldwide result in the existence of different English varieties (Kachru, 1992; Schneider, 2003, 2007).

This project is significant because it explores CS, one of the effects of language contact in multilingual communities such as Zimbabwe. In this study, CS between English and Shona was found to occur, with the majority of instances being English as the Matrix Language (ML) and Shona as the Embedded Language (EL) (see Myers-Scotton, 1993b). However, this study will investigate both directions of distribution, including both English and Shona as the ML. The aims of this study are to identify the different types of CS, to ascertain the positions where CS occurs in different discourses in the ZimE corpus, and to determine their frequencies within the grammatical perspective. In addition, this study analyses the functions of CS in the ZimE corpus to ascertain the frequencies of the various functions in terms of the socio-pragmatic approach, a crucial step in understanding the linguistic-, cognitive-, and sociocultural dimensions of bilingualism in Zimbabwe.

The foci of this study, firstly, included the identification of the positions of different types of CS in such discourses as are included in the corpus and secondly the study of their frequencies which can have implications in five different specialization areas. These five areas include language contact, sociolinguistics – specifically understanding bilingual proficiency, informing psycholinguistic models of bilingualism (Green and Wei, 2014), language acquisition and evolution, and informing theoretical models underlined by real-world bilingual speech patterns (Bullock and Toribio, 2009).

The following research questions guided the research. Questions 1 to 4 were analyzed within the grammatical perspective.

1. What are the types of processes with regard to code-switching and borrowing that are found in the ZimE corpus?
2. What are the specific types of code-switching that are found in the ZimE corpus?
3. What is the frequency of the different types of code-switching in the ZimE corpus?
4. What is the frequency of the different types of code-switching for spoken and written registers in the ZimE corpus?
5. How are CS and borrowed forms used and distinguished in the ZimE corpus?
6. What are the possible functions associated with the code-switching which occurs in the ZimE corpus, specifically within the socio-pragmatic perspective?
7. What is the frequency of different functions of CS in the ZimE corpus?

2. LANGUAGE USE IN ZIMBABWE

Similar to many countries in Africa, diversity is a characteristic of language use in Zimbabwe due to the variety of languages that are spoken, thereby leading to language contact (Myers-Scotton, 1993a). Barwe, Chewa, English, Kalanga, Khoisan, Nambya, Ndau, Ndebele, Shangani, Shona, Sotho, Tonga, Tswana, Venda, Xhosa, and Zimbabwean Sign Language are officially recognized by the Constitution of Zimbabwe Amendment Act (Number 20) of 2013. Despite the recognition of 16 official languages, only English, which was imposed during British colonialism, still dominates in teaching and learning, government, industry and trade, parliamentary business, and the print media, whilst indigenous languages are marginalized and spoken mainly at home (Kadenge 2010). It is inevitable that the Zimbabwean linguistic environment where English and indigenous languages are used, leads to language contact, which manifests in CS, code-mixing, and borrowing, among other features. CS and borrowing are of interest to this study, both of which are discussed in section 3.

Whilst this study is concerned with CS involving English as the ML and Shona as the EL in the majority of occurrences, English is used as L1 by only 5 % of the population in Zimbabwe (Fitzmaurice, 2020). English is furthermore acquired as an L2 at school by most indigenous L1 language speakers in Zimbabwe (Kadenge and Mugari, 2015). On the other hand, the term "Shona" came into existence after Doke (1931) studied different dialects (Karanga, Korekore, Manyika, Ndau, and Zezuru) and noted that they are mutually intelligible and can be referred to under the umbrella term "Shona". This identification led to the unification, standardization and development of Shona that is still ongoing. In terms of categorization, Shona belongs to Zone S10 in Guthrie's (1948) classification of Bantu languages and is a tonal and agglutinative language. Such languages use strings of morphemes in a linear fashion to create new words (Fortune, 1984). Shona has 21 noun classes (see Table 1 for some of the noun classes) and its basic word order follows a Subject-Verb-Object (SVO) pattern. The addition of prefixes or suffixes to existing words to create new words is also a characteristic of Shona derivational morphology (Mkanganwi, 2002).

Table 1
Noun-class prefixes in Shona (Sources: Fortune, 1984: 31f; Déchaine et al., 2014: 20).

Descriptive semantic features	Noun class	Noun class prefix	Singular/ Plural	Examples	
Human	1	*mu-*	singular	*mu-nhu*	'person'
	2	*va-*	Plural	*va-nhu*	'people'
Solid, Extended	3	*mu-*	singular	*mu-ti*	'tree'
	4	*mi-*	Plural	*mi-ti*	'trees'
Solid, Non-extended	5	VOICE-	singular	VOICE-*gore*	'cloud'
	6*	*ma-*	Plural	*ma-kore*	'clouds'
Artifact	7	*chi-*	singular	*chi-rongo*	'waterpot'
	8	*zvi-*	Plural	*zvi-rongo*	'waterpots'
Animal	9	N-	singular	N-*shumba*	'lion'
	10*	N-	Plural	N-*shumba*	'shumba'
Outline, Extended	11	*ru-*	singular	*ru-kova*	'stream'
	10*	N-	Plural	N-*hova*	'streams'
Small entity	12	*ka-*	singular	*ka-mbambaira*	'small sweet potato'
	13	*tu-*	Plural	*tu-mbambaira*	'small sweet potatoes'
Solid, Differentiated	14	*Xu-*	singular	*u-swa*	'grass'
	6*	*ma-*	Plural	*ma-uswa*	'grasses'
Infinitives	15	*ku-*		*ku-enda*	'to go'
Locatives (*on/ at*)	16	*pa-*		*pa-chipatara*	'at the hospital'
Locatives (in/ inside)	17	*mu-*		*mu-poto*	'in the pot'
Small, Tiny	19	*svi-*	singular	*svi-sikana*	'tiny girl'
Augmentative	21	*zi-*	singular	*zi-munhu*	'big person'
	6*	*ma-*	Plural	*ma-zi-vanhu*	'big people'

* An asterisk placed next to a noun class means that the noun class is multifunctional.

3. CODE-SWITCHING: DEFINITIONS AND THEORETICAL PERSPECTIVES

CS, a characteristic of bilingualism and multilingualism, occurs as an increasing global phenomenon due to contact between different languages in speech communities (Bullock and Toribio, 2009). CS is variably defined depending on the approach taken, as evidenced by some of the definitions found in literature. For instance, Wardhaugh (1992) refers to CS as the alternation between the two systems used by people to communicate, while Myers-Scotton (1993a: 4) defines CS as "the selection by bilinguals or multilinguals of forms from an embedded language (or languages) in utterances of a matrix language during the same conversation". Another definition is that CS is "the alternation of two languages within a single discourse, sentence or constituent" (Poplack, 1980: 583). Although controversy surrounds what constitutes CS, this study uses the term "code-switching" to refer to switches that occur as single morphemes, single word switches or longer stretches of words (see Myers-Scotton, 1993b).

Distinctions have been made between intrasentential CS, intersentential CS and extrasentential CS (tag switches). Intrasentential CS is the switching that "occurs within the same sentence, from single morpheme to clause level" (Myers-Scotton, 1993a: 4). Intersentential CS involves switching that occurs at sentence boundaries, and extrasentential CS or tag switching occurs when an exclamation or a tag element "from one language is inserted into an otherwise monolingual discourse of another language" (Koban, 2013: 1175). Koban further notes that their syntactic nature enables tag switches to be placed in different places without violating the syntactic order of sentences. The distinction between intrasentential CS and extrasentential CS is very nuanced as in both CS types, a single morpheme or phrase from the embedded language is inserted into the ML. However, extrasentential or tag switching involves incorporating discourse markers, interjections, or tags from one language into another where these tags are elements that are not reliant on the sentence's syntax. Lastly, it is important to note that Muysken's (2000) typology explains code-mixing in terms of insertion, alternation and congruent lexicalization.

Diversity characterizes the study of CS because researchers approach the topic from different perspectives such as grammatical- (e.g. Myers-Scotton, 1993b), socio-pragmatic- (e.g. Gimode and Barnes, 2015), sociolinguistic- (e.g. Blom and Gumperz, 1972; Gumperz, 1982; Myers-Scotton, 1993a), and psycholinguistic perspectives (e.g. Grosjean, 1995). Pertinent to this study are the grammatical and socio-pragmatic perspectives. The grammatical approach focuses on determining the structures in CS and where in different discourses CS occurs, thus illuminating the resourcefulness of speakers (e.g. Myers-Scotton, 1993b). According to Nel (2012), the grammatical characteristics of CS can be

explained in terms of three different approaches, namely the variationist-, generativist- and production approaches. The variationist- and generativist approaches formed the basis for the production-based approach of the Matrix Language Frame (MLF) model. This model consisting of the Matrix Language Hypothesis, aims to explain the sentence structure in intrasentential CS (Myers-Scotton, 1993b, 1997). The asymmetrical nature of the MLF model is evident in the fact that it designates the dominant language as that which dictates morpheme order and builds the abstract grammatical structure of the sentence as the ML, whilst the EL supplies "either content morphemes in mixed constituents or EL phrase-level constituents (EL islands)" in the mixed segments (Myers-Scotton and Jake, 2009).

The asymmetrical relationship between the ML and the EL produces classic CS.[1] The MLF model also distinguishes between system and content morphemes. In this distinction, content morphemes (CMs) assign or receive thematic roles whilst system morphemes (SMs) are words such as affixes and function words with little lexical meaning. In this model, clitics and determiners are classified as SMs (Myers-Scotton and Jake, 2009). Furthermore, the Morpheme Order Principle is used to designate the ML as the language to be followed in terms of how morphemes are arranged, while the System Morpheme Principle states that the source of the SMs in a bilingual constituent is the ML.

The MLF model has been revised and augmented several times. For example, the 4-M model has been added in order (i) to classify different morphemes, (ii) to highlight their syntactic functions, and (iii) to outline the processes involved in accessing different morphemes at different language production levels (Myers-Scotton and Jake 2001). As with the MLF model, the 4-M model upholds the distinction between CMs and SMs but further divides these categories into early SMs and late SMs. The two types of late SMs are referred to as bridges, made up of a single allomorph, and outsiders, which have more than one allomorph (Myers-Scotton and Jake 2001). Due to their salience in the mental lexicon, early SMs, together with their content morpheme heads, function as primary units in constructions such as noun phrases, verb phrases and adjective phrases (Myers-Scotton and Jake, 2009). The 4-M model postulates that, in contrast to early SMs, which predominantly frame the semantic structure, late SMs frame syntactic structure, and form the bonds that are responsible for clause coherence. The MLF model, together with the 4-M model are pertinent to our analysis because they were used as the basis for analyzing CS patterns to determine the types, frequencies and functions of CS in the ZimE corpus. Table 1 provides a summary of the types of morphemes which can be used in English-Shona CS.

3.1. Functions of code-switching

Previous research has documented different functions associated with CS. These functions include, first, switching at a time of vagueness and uncertainty (De Klerk, 2006). This type of switching is usually preceded by both filled and unfilled pauses and speakers switch when they have difficulties finding words for concepts in the ML. Another characteristic of switching at a time of vagueness or uncertainty is repetition. Second, CS occurs in order to make a point clear, to clarify, and to elaborate. Stressing or emphasizing a point is a third function of CS (Poplack, 1980). Speakers may, as a fourth purpose, also codeswitch for repetition or reiteration to make a point clearer (De Klerk, 2006), and for the purpose of asking questions (Rose and van Dulm, 2006: 9). A fifth purpose is that of identity marking, which involves strengthening or affirming cultural identity (De Klerk, 2006). Such marking occurs when speakers from the same ethnic or cultural group switch to show a shared identity. Sixth, discourse marking is one of the prevalent functions of CS (De Klerk, 2006) where speakers often insert discourse markers either at the beginning, middle or end of a sentence as a way of engaging or encouraging communication.

Seventh, the emotive function of CS is shown when interlocutors switch to communicate their moods and feelings. Due to the differences in the way languages use words to express emotions, speakers may switch to another language in order to better convey their feelings (De Klerk, 2006). Lastly, CS is also used as an inclusion tool to accommodate other speakers. Finlayson and Slabbert (1997) illustrated this function of inclusion in a study of CS involving 42 participants from Soweto, South Africa. Results showed CS being used to include other speakers, especially in the multilingual community of Soweto where different South African languages are spoken. A detailed qualitative analysis of the functions of CS and examples are provided in section 6.4.

3.2. Distinguishing between code-switching and borrowing

Distinguishing between borrowing and CS is not clear-cut due to the two concepts falling on different points of a continuum while also incorporating different types of the two phenomena (Myers-Scotton, 1997). Borrowing involves adopting words or phrases from one language to another, especially to fill lexical gaps in the borrowing language. On the

[1] "CS in which empirical evidence shows that abstract grammatical structure within a clause comes from only one of the participating languages" is defined as classic CS (Myers-Scotton and Jake 2009: 337).

other hand, language alternation occurs in CS which is not meant to fill lexical gaps, so that the borrowed words remain unassimilated into the borrowing language (Kamwangamalu, 1989). L2 English varieties spoken in Africa show a large number of borrowed words from indigenous languages into English, especially terms for indigenous foods and the flora and fauna (Schneider, 2003, 2007). Another difference is that borrowing can occur in the speech of both monolinguals and bilinguals, whilst CS is confined to bilinguals or multilinguals with varied linguistic proficiency levels (Gumperz, 1982; Kamwangamalu, 1999). Unlike in borrowing, the phonological and morphological features of the switched items sometimes do not change in CS (Sridhar and Sridhar, 1980).

Observations have also been made about the different types of borrowing, among them lexical borrowing, that is, when words are borrowed from another language (Haspelmath and Tadmor, 2009), and loanword adaptations, in which the morphological and phonological integration of loanwords has occurred to fit the receiving language (Kamwangamalu, 1999). Regarding lexical borrowing, ZimE has loaned and incorporated words such as *sadza [maize meal]*, and names for indigenous trees like *muonde [fig tree]* (Kadenge, 2012). Examples of loanword adaptations from English to Shona include *chikoro [school]* and *dhokotera [doctor]* (Ngara, 1982).

Since divergent views are abounding in the literature and considering that there is no explicit boundary between CS and borrowing, this study used predictability, a criterion which encompasses acceptance of an element by a speech community (Muysken, 2000) to distinguish between the two. The predictability criterion assumes that the vocabulary of monolingual speakers of the receiving language contains a list of loanwords or lexical borrowings whilst this is not the case for CS (Myers-Scotton, 2002). It is necessary to keep in mind that lexical borrowings do not always cover lexical gaps in the target language. Many lexical borrowings can be incorporated even if an equivalent form exists in the target language. Both lexical forms may coexist in the target language for a long time (Grant, 2015). Therefore, items in this study were classified as borrowings if they were considered to be filling lexical gaps (e.g. Shona terms for indigenous foods, traditional games, cultural practices, and traditional ceremonies). In addition, instances where the equivalent forms existed were classified as lexical borrowings. On the other hand, instances where there was language alternation not meant to fill lexical gaps and where the borrowed words remained unassimilated and not integrated were regarded as CS (Sridhar and Sridhar, 1980; Kamwangamalu, 1989; Bullock and Toribio, 2009).

4. PREVIOUS RESEARCH ON ENGLISH-SHONA CODE-SWITCHING

Previous research on English-Shona CS, where English is the ML and the dominant language and Shona is the EL (see Myers-Scotton, 1993b, 1997), has helped shed light on its nature, functions and characteristics from different perspectives. For example, Mukenge and Chimbarange (2012) used a sociolinguistic approach to examine CS as a strategy for communication in the film *Yellow Card*. The study reported on the complexity of CS in the expression of a speaker's intentions. Marungudzi (2014) also utilized a sociolinguistic approach in the analysis of English-Shona CS in Zimbabwean classrooms. Results from the study point towards the creative use of CS to assist teachers to successfully transmit content, and as a classroom management strategy.

Using a sociolinguistic approach, Nyamayedenga (2022) examined how multilingualism affects the implementation of the communicative language teaching method due to CS. Nyamayedenga (2022) emphasized the importance of CS as a powerful technique that enhances learning. Gotosa, Rwodzi and Mhlanga (2013) studied CS in learning from a socio-pragmatic perspective. The authors noted that although English is supposed to be used in the classroom, CS is the norm and should be accepted due to the functions it serves in the classroom. In another socio-pragmatic study, Viriri and Viriri (2013) used a case study of schools in the Buhera South district in Zimbabwe and noted that, despite the expectation that English would be used, teachers and learners often switched between English and Shona to clarify certain points during the learning process.

Although the abovementioned studies have focused on English-Shona CS, the subject has not been adequately studied from a grammatical approach and through the use of a bigger corpus in order to better explore and understand the frequencies and functions of CS. The analysis of CS in the ZimE corpus, as conducted in the current study, will have implications for the natural language processing (NLP) and machine learning fields, some of the areas that have gained scholarly interest on CS data (Aguirre et al., 2022). NLP tasks are data driven, requiring CS texts to train and test applications in tasks such as automatic speech recognition (ASR) and text-based applications (Aguirre et al., 2022). It is necessary to collect data that can be used to study CS, especially in NLP where there is a need to develop language models that can accurately handle and process code-switched language.

5. RESEARCH METHOD

The ZimE corpus consisting of 356,007 words was used for this study. The corpus is made up of 206,007 words collected by the researcher (Chapwanya, 2022), while the additional 150,000 words were taken from Marungudzi's

(2016) corpus. The overall number of words in the ZimE corpus includes English words and Shona words which are incorporated as a result of CS. The motivation for collecting a corpus stemming exclusively from the English of Shona speakers instead of other indigenous languages of Zimbabwe such as Barwe, Chewa, Kalanga, Khoisan, Nambya, Ndau, Ndebele, Shangani, Sotho, Tonga, Tswana, Venda and Xhosa was that the languages are significantly different and may have influenced English differently (see De Klerk, 2006).

The 206,007 words compiled by the researcher were drawn from various participants and represented different registers, as follows: 20 for private dialogues, 45 for private semi-scripted dialogues, 5 for business letters. Online samples of editorials and newspaper reportage were taken from 10 and 14 writers, respectively. In total, 94 participants were included in the corpus. For Marungudzi's (2016) corpus which supplied 150,000 words, information on the number of participants is not available. All participants who provided both spoken and written texts signed informed consent forms which explained participants' rights and roles. For preservation of anonymity and confidentiality during data processing and reporting, pseudonyms were used in the representation of participants. All spoken and written data collected, and the information provided by or about the participants for this study was stored on a password protected USB flash drive and conformed to the POPIA act of South Africa.

Different instruments that included a biographical and language background questionnaire and semi-structured interviews were used to determine whether participants could converse in English. The interviews and the questionnaire were valuable because they were used for gathering linguistic proficiency data in order to aid the discussion on the language backgrounds of participants. Informed consent forms were used to gather data for spoken texts such as private dialogues, public scripted and unscripted monologues, and written texts, namely social letters. Participants provided their biographical information and information about their linguistic backgrounds to determine their suitability to participate in the study. A smartphone and a digital recording device were used to record private dialogues, public scripted monologues and unscripted monologues for a duration of 15 to 30 min. The recordings were labelled to identify each participant and pseudonyms were used to preserve the confidentiality of all participants' information and anonymity in reporting.

Spoken and written samples containing at least 2,000 words per text sample were gathered in both rural and urban settings in and around Gweru, Harare, Masvingo, and Mutare. Speech samples included private dialogues (10 texts) and private semi-scripted dialogues (45 texts). Written samples included business letters (5 texts), editorials (10 texts), and newspaper reportage (14 texts). Online texts from *The Herald* and *The Sunday Mail* newspapers were used for newspaper reportage and editorials. The texts from which Marungudzi (2016) drew samples for the 150,000 word corpus are two public scripted monologues, 47 public dialogues, eight creative writing texts, one popular writing text, one academic writing sample, one newspaper reportage, two social letters, three business letters and one sample for editorials.

The non-probability sampling technique involving convenience, purposive and snowball sampling methods was used (see Leedy, Ormrod and Johnson, 2021). Non-probability sampling encompasses various strategies that try to attain a reasonably representative sample using resources that are readily available to the researcher (Dörnyei, 2010). Participants are selected based on the nature and objectives of the study. Diversity characterized participants' socio-cultural and linguistic backgrounds. The criteria for inclusion in the study were that participants should be 18 years or older, be Shona L1 speakers who use English as an L2, have attended formal education for at least 10 years, and have the ability to answer the interview questions in English. Even though this study did not measure language proficiency, insights about the levels were gained using biographical and language background questionnaires that were completed by participants. From the questionnaires and interview questions administered, assumptions were made that learners who had at least 10 years of formal education would have been sufficiently exposed to English and could communicate in English at a level to be considered bilingual due to the influence and use of English in the Zimbabwean education system (see Bullock and Toribio, 2009). For this reason, participants' linguistic proficiency levels were varied.

As was done by Marungudzi (2016), the international corpus of English (ICE) design and annotation conventions were followed in the current study (see Nelson, 2002a, 2002b). This enabled uniformity in the markup symbols used, thereby making it easy to analyze the corpus when Marungudzi's texts were combined with texts compiled by the researchers. Regarding CS, different corpora compiled following the ICE protocol, including the current study, annotate language alternation with the markup symbol <indig> </indig> (Nelson, 2002a; 2002b). The word count for each text type in the ZimE corpus is presented in Table 2.

This study used a corpus-based approach which involves the gathering of observational data to observe and analyze naturally occurring language in order to identify and quantify linguistic features (Gries, 2006). The approach combines quantitative and qualitative methods. The quantitative approach was used to obtain statistical data such as frequencies of the types of CS in the corpus and their functions. The qualitative analysis involved establishing the nature and functions of the various occurrences of CS. By evaluating the frequencies and types, more insight was gained about how CS

Table 2
Word count for text types in the ZimE corpus.

Register	Text type	Number of texts	Number of words
Spoken	Private dialogues	10	25,551
	Private semi-scripted dialogues	45	106,410
	Public scripted monologues	2	4,193
	Public dialogues	47	98,760
Written	Creative writing	8	17,890
	Popular writing	1	2,759
	Academic writing: Examination	1	2,077
	Editorials	11	26,342
	Newspaper reportage	15	48,939
	Social letters	2	5,008
	Business letters	8	18,078
Total		150	356,007

is used, and which aspects are predominant in the Zimbabwean language context, which may have implications for language policies in Zimbabwe.

The ZimE corpus was uploaded to Sketch Engine Tools software (www.sketchengine.eu/) or automatic part-of-speech (POS) tagging and analysis of CS between English and Shona. In the wordlist tool, the tagset <indig> </indig> which shows parts where switches from English to Shona were searched for. This search produced a concordance list showing the key words in context (KWIC). Clicking on every instance where <indig> was used generated the context where the tagset occurred. Each context was expanded to enable determination of whether English was the ML or Shona was the ML for the analysis. It should be noted that although automatic POS tagging was done in Sketch Engine, this was used sparingly. Since some POS tags were inaccurate, manual annotation was done using the Sketch Engine Manual annotation (Skema) tool. The adjustments were saved for later analysis. For semantic and pragmatic interpretations of the contexts where switches occurred in order to determine the functions of CS, the Skema tool was used. Concordance lines where the switches occurred were searched for and annotated according to their primary functions. This manual annotation enabled instances of borrowing to be distinguished from CS.

Although there is no explicit distinction between CS and borrowing, contextual cues within the concordance lines and the descriptions (given by Sridhar and Sridhar 1980; Kamwangamalu, 1989; Kadenge, 2012) enabled these distinctions to be made. The frequencies of both the types and functions of CS in the ZimE corpus were normalized per 100,000 words and rounded off to the nearest whole number (as reported on in the results sections). In the examples given, all Shona words were italicized, together with their translations (which were given in square brackets).

5.1. What constitutes code-switching data in this study

CS data in the current study included intrasentential CS, intersentential CS and extrasentential CS and included multi-level items such as morphemes, words and clauses as part of the analysis (see Myers-Scotton, 1993b). Words which had bound morphemes such as prefixes and suffixes were considered CS data in this study. For the purpose of distinction, intrasentential CS occurs inside a sentence, as illustrated in example (1). On the other hand, intersentential CS involves alternating from one language to another at sentence boundaries, as shown in example (2) where the first sentence is in English, and intersentential CS to Shona occurs in the second sentence.

(1) Ok, I will come. I need clothes for ***vazukuru vangu*** *[my grandchildren].* (Private dialogues 10)

(2) <$LL>: I would like to share with you Genesis one verse one. ***Pakutanga Mwari vakasika denga nenyika*** *[In the beginning God created heaven and earth].* (Private semi-scripted dialogue 38)

Intrasentential CS can be subdivided into (i) single word switches (see example 3), (ii) multi-word switches in different places not adjacent to each other (see example 4 where *handiti [is that so]* is also characterized as a tag switch), (iii) multi-word switches in the same place and adjacent positions (islands) (example 5), and (iv) below word level switches (switching that involves bound morphemes, as in example 6).

(3) <$P>: ***Isusu*** *[Us]* like us men then we say uh we are in charge. (Private semi-scripted dialogue 16)
(4) <$D>: I believe last time *pakaita* congress *ku*Gweru *[when congress happened in Gweru]* that issue was raised even by the President ***handiti*** *[is that so?]* (Public dialogue 17)

EL islands refer to speech segments from a less dominant language (the EL) inserted within a sentence containing a dominant language, referred to as the ML (Myers-Scotton, 2002, 2005). In EL islands, the constituents are exclusively in the EL and follow the EL grammar, as shown in example (5) where *pahongo yeminda [during the fields war]* is the Shona EL island. Example (5) consists of the ML constituents: *I got a farm from the government um I think it was ten years ago* and the EL island *pahondo yeminda*. From the analysis above, it is important to point out that intersentential CS typically happens at clause boundaries, whereas EL islands consist of multi-word phrases in one language embedded within a single clause. Example (6) illustrates below word level switches (switching that involves bound morphemes such as *ma-*).

(5) <$Y>: I got a farm from the government um I think it was ten years ago ***pahondo yeminda*** *[during the fields war].* (Private semi-scripted dialogue 25)
(6) <$G>: And the some organisations donated ***ma***-tablets, those solutions to people (Private semi-scripted dialogue 7)

The ML change from English to Shona occurred when the sentence structure followed the Shona language grammatical conventions (see Myers-Scotton, 2005). Thus, if the tense, word order and agreement were in line with Shona rules, the ML had changed (Myers-Scotton, 1993b). In addition, an analysis of clause boundaries (to check which language provided the morphosyntactic frame) assisted in determining whether the ML had changed from English to Shona. Example (7) shows a change from English as the ML to Shona as the ML.

(7) <$A>: Yes good morning <,> you are live on radio.
<$D>: ***Ndoda kubvunza chiremba kuti ndine mwana wanguwoka akaa***<,,> ***akayamwira*** **for five months** ***ndisingazivi kuti ndine pamuiri handizive kuti zvinomu*****affecta kudii**<,,>. *[I want to ask the doctor that I have a child who <,,> who drank milk for five months while I didn't know that was pregnant. I don't know how this will affect the child <,,>].* (Public dialogues 3)

6. RESULTS

6.1. Frequencies of processes and switching types

This section provides quantitative answers to research questions 1 to 4, focusing on (i) the processes of language alternation (research question 1), (ii) the identification of the different types of CS (research question 2), and (iii) the frequencies with which CS (research question 3) was found in the ZimE corpus (overall), as outlined in Table 3. The study focuses on occurrences of both CS and borrowing in the corpus. To be able to draw conclusions on CS, both CS and borrowed items were included in the analysis in Table 3. In the ZimE corpus, there were instances where the ML chan-

Table 3
Frequencies of switching according to the ML (normalized per 100,000 words).

Types of processes and switches		English matrix language		Shona matrix language	
		Raw frequency	Normalized frequency	Raw frequency	Normalized frequency
CS	Intersentential CS	135	38	6	2
	Intrasentential CS	311	87	35	10
	Extrasentential CS	17	5	3	1
Borrowing	Lexical borrowing	66	19	5	1
	Loanword adaptations	8	2	0	0
Total (CS and borrowing)		537	151	49	14

ged from English to Shona. This was observed especially in the private and public dialogues where participants showed a preference for Shona instead of English. Some participants requested to be allowed to speak in Shona because they felt comfortable expressing themselves in their mother tongue. Another reason for switching the ML from English to Shona could be due to the topics under discussion. For instance, there was switching during discussions about medical issues, legal matters and politics. Results in Figs. 1 and 2 reflect such cases.

In total, there were 537 switches where English was the ML in the ZimE corpus, translating to 151 normalized occurrences. Intrasentential CS was more prominent, occurring 87 times, whilst intersentential CS was attested 38 times. Interestingly, lexical borrowing was the third largest type of switching and occurred 19 times. Extrasentential CS was less prevalent, only occurring five times. Loanword adaptations were attested least in the corpus, with only two instances being identified. With regard to switching where Shona was the ML, intrasentential CS had the highest frequency of 10 occurrences, followed by intersentential CS with two occurrences. Extrasentential CS and lexical borrowing were both attested only once.

Fig. 1 illustrates the results of a micro-level analysis of intrasentential CS. The intrasentential CS types were divided into different categories in order to gain better insights into the CS patterns within sentences. In this study, bound morphemes which included below word level switches (such as prefixes and suffixes) were classified as intrasentential CS.

The remarkable nature of the data in Fig. 1 lies in the fact that multi-word switches in the same place (islands) occurred 53 times in English ML utterances compared to 14 times in Shona ML utterances (as normalized occurrences). Most multi-word switches in the same place were attested in the public dialogues. Since this genre included multiple speakers, the switches may have been a way to accommodate other speakers. This is discussed further in section 7. Another noteworthy observation is that below word-level switches, where English words were affixed with Shona bound morphemes, were prevalent in English ML sentences. The Shona prefixes and suffixes were added to English words, as in ***ma***-*symptoms*, ***ndi***-*matron,* ***pa***-*chest*, ***mu***-*constitution*, and ***ye***-*constitution* (see further discussion in sections 6.2.3 and 6.2.4). These switches were considered as intrasentential CS involving below word level switches. The normalized frequency for single word switches which included extrasentential CS in English ML switches was 23 compared to three in Shona ML switches. The single word switches and extrasentential CS were combined for ease of analysis since the two types of switches share similarities (see Koban, 2013). The analysis revealed below word level switches, instances where affixation occurred to a single English word in a sentence (attested three times in English ML switches as opposed to only once in Shona ML switches). The majority of affixation was done on multi-word switches in different places in an utterance.

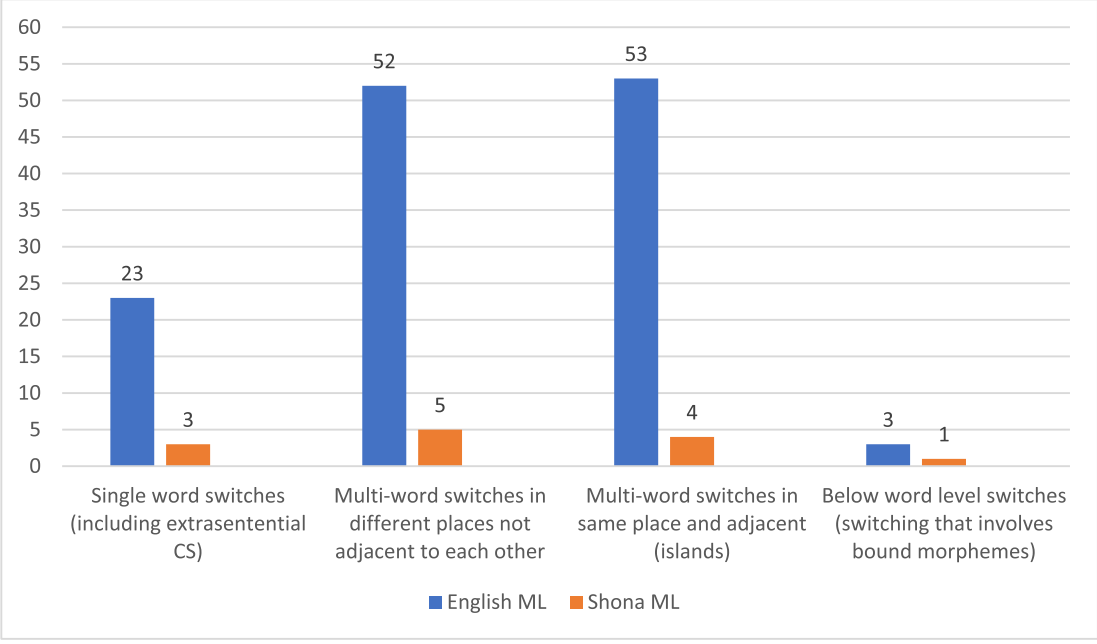

Fig. 1. Frequency of different intrasentential code-switching types (normalized per 100,000 words).

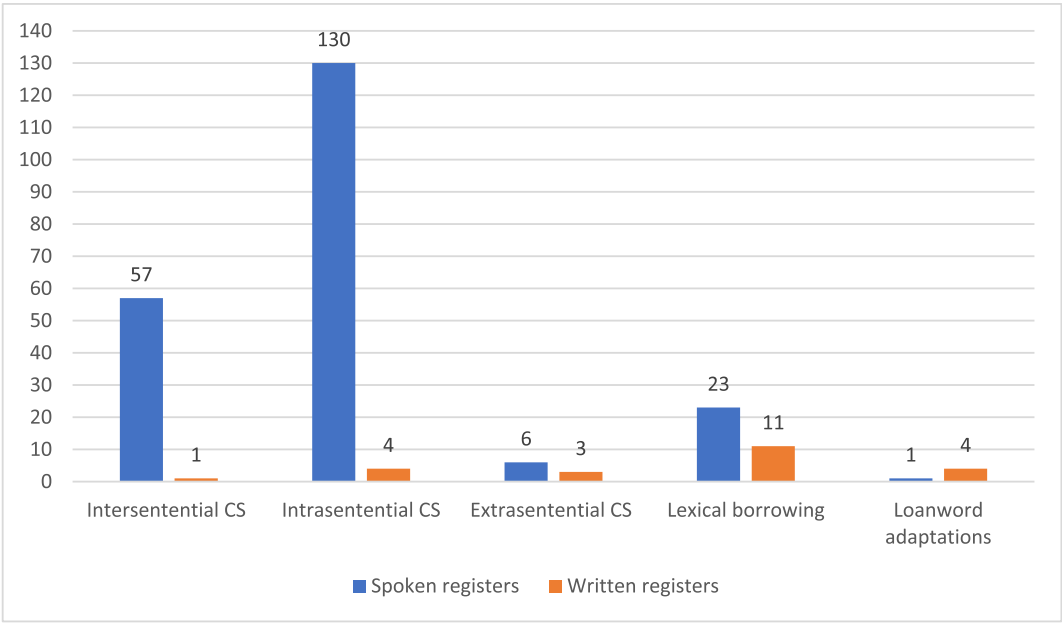

Fig. 2. Frequencies of switches in spoken and written registers (normalized per 100,000 words).

This section also focused on the frequency of the different types of code-switching for spoken and written registers (research question 4) in the ZimE corpus within the grammatical perspective, as illustrated in Fig. 2.

Corpus analyses of the frequencies of CS types in spoken registers showed that intrasentential CS recorded the highest frequency of 130 occurrences. This is due in part to the inclusion of CS that involves bound morphemes as part of the intrasentential switches (outlined in Figs. 1 and 2). The bound morphemes included below word level switches that were affixed with Shona prefixes and suffixes. Intersentential CS was the second highest CS type attested in the ZimE corpus (with 57 occurrences). Items that were regarded as lexical borrowing occurred 23 times, whilst loanword adaptations were the least frequent in the corpus, occurring only once. The picture is different regarding the written register where lexical borrowing accounted for the highest frequency with 11 occurrences. Intrasentential CS and loanword adaptations occupied second spot (with four occurrences). Overall, there were more instances of CS in the spoken registers (217 occurrences) compared to the written registers (23 occurrences). These results confirm previous findings which reported the high prevalence of CS in spoken registers due to their spontaneous nature (see Barasa, 2016).

6.2. Evidence of code-switching and borrowing in the ZimE corpus

This section will thus focus on answering research question 5 to qualitatively analyze the different alternation processes of CS and borrowing which occur in the corpus. Firstly, all instances of CS will be examined, followed by a discussion of the cases of borrowing found in the corpus. Instances of CS occurring in the corpus can be characterized by plural morpheme doubling on nouns, the use of English nouns with Shona prefixes and suffixes, and the use of English verbs with Shona prefixes and suffixes.

6.2.1. Plural morpheme doubling on nouns

Plural morphemes are doubled most often in CS (Myers-Scotton, 1993b). The process involves "the occurrence of an EL system morpheme with its ML counterpart in a mixed constituent" (Myers-Scotton, 1997: 102) That is, the EL plural affix is accessed along with the EL noun before the ML fills in the frame with its version of the plural affix. For instance, in example (8), the Shona plural prefix for noun class 6 is added in the sentence where Shona is the ML and English is the EL.

(8) But ***ma***-day-***s*** *ano aya handisi kumuona [But these days I don't see him]* (Myers-Scotton 1997: 111).

On 34 occasions, plural morpheme doubling occurred with the addition of the Shona plural prefix for noun class 6 (*ma-*) to the words that already have the English plural marker suffix (*-s*). Besides the majority of plural morpheme doubling occurring in English ML phrases, it also occurred in Shona ML phrases, as illustrated in example (8). Examples of words and phrases with plural morpheme doubling include *ku-**ma**-union**s** [at unions]*, *ne-**ma**-individual**s** [and individuals]*, *e-**ma**-employee**s** [of employees]*, *se-**ma**-citizen-**s** [as citizens]*, *ye-**ma**-possession**s** [of possessions]*, *pa-**ma**-right**s** [with regard to* rights*]*, and *dze-**ma**-property right**s** [of property rights].* In the aforementioned examples, there is also pre-prefixing whereby "one or more prefixes of other noun classes are inflected onto a stem that already carries a prefix denoting the primary specimen of that class" (Mashiri, 2002: 252). The pre-prefixes include *ku-*, *ne-*, *pa-*, *e-*, *se-*, *ye-*, *and dze-*. The normalized frequency of plural morpheme doubling in the ZimE corpus is 10 per 100,000 words. Examples (9) and (10) illustrate plural morpheme doubling in context:

(9) <$G>: Good good <#> E e you know what my brother Napster. Problem *yacho inotangira pama*confidentiality *anenge achingotaurwa iwayaya* <,> *zvokuti **ma***-disclosure-***s** [the problem starts at confidentialities that are talked about tha the disclosures]* are not adequate but obviously *chinoitika ndechekutika [what happens is]*, if we need to take a holistic approach, even the general employees as well <,> ***ma***-salarie-***s** avo [their salaries]* should be reasonable. (Public dialogue 25)

(10) <$O>: The teachers also don't have much knowledge on this new curriculum thing. <#>So, it's affecting us so much. <#>We are not getting enough knowledge about it. <#>And I think it it requires a bit more expensive things like ***ma***-computer-***s***, laptop-***s***, cell phone-***s***. (Private dialogues 8)

What is interesting about example (10) is that participant <$O> uses double plural morphemes in ***ma**-computer-**s*** and then goes on to use English plural marker (*-s*) for the words *laptop-**s*** and *cell phone-**s***. This could be a result of self-correction whereby a speaker becomes aware of the linguistic norms and corrects themselves in real time. In this instance, the speaker code-switched using ***ma***- which belongs to noun class 6. Since *computers*, *laptops*, and *cell phones* are content morphemes which assign or receive thematic roles, self-correction occurred in order to preserve sentence coherence, as suggested in the 4-M model (Myers-Scotton and Jake, 2009). The 4-M model postulates that early SMs usually conform to the ML sentence structure. This was the case with regard to plural morpheme doubling in the ZimE corpus. In all instances where English was the ML, the English plural marker (*-s*), an outsider late system morpheme, was added to English content morphemes. This is because early SMs add detail to the meaning of their content morpheme heads whilst late SMs assist with syntactic structure. The high frequency of *ma-,* a Shona plural morpheme, can be linked to the 4-M model's prediction that during CS, all morphemes added to a sentence conform to both the semantic and syntactic structures of a sentence in order to achieve clause coherence. Even though intrasentential CS occurs, the inclusion of *ma-* does not violate the syntactic structure of English, the ML. This finding shows morphosyntactic alignment at play, with speakers trying to follow the grammatical rules of both languages.

Plural morpheme doubling also occurred when the ML changed from English to Shona, as highlighted in example (11).

(11) <$C>: ***Ma***-union-***s** havafaniri kunge vachitaura nezvemitengo yedonj*e. (Public dialogue 22) [*The unions should not talk about cotton prices*].

6.2.2. English nouns with Shona prefixes and suffixes

An interesting observation was made regarding the use of Shona prefixes and suffixes in English words and phrases. For instance, prefixes were used with proper nouns which represent individuals, places, things and organizations, such as are outlined in examples (12) and (13)**.**

(12) <$D> Businesses are low. *Vamwe vari kutovhara* so ***pa***-Christmas *[Others are closing at Christmas]* and I don't think people are going to come back next year <,> *kumabasa avo [to their jobs]*. (Public dialogues 47)

(13) This sort of common often meant that gangs of *matsotsi* ***e***-Harare *[Harare thieves]* were out prowling thro the ramshackle labyrinth of Jo'burg Lines where she lived with. (Creative writing 2)

Shona prefixes were also frequently used with common nouns in the ZimE corpus. Example (14) shows such usage where the noun *stage* was prefixed with *pa-* and example (15) shows how prefixes *we-* and *mu-* were added to the nouns *council* and *government.*

(14) <$D>: Uh every step of the way *vaiuya vachitidzikamisa hana vachitaura kuti tasvika* ***pa-***stage *yakati*. *[they came to calm us down and told us the stage they were at].* (Private dialogues 29)

(15) <$G>: You can't have a situation *yekuti [where]* nurse ***we-mu-***council *[in council]* earns two thousand nurse ***we-mu-***government *[in government]* earns four hundred.

The use of Shona suffixes with common nouns was attested three times in the words *question-**ka***, *board-**ka***, and *a concerned voter-**ka***.[2] The frequency of Shona prefixes and suffixes in the ZimE corpus was 28 occurrences per 100,000 words. Example (16) illustrates Shona prefix use, whilst example (17) shows Shona suffix use in English words and phrases.

(16) While growing up, we always knew that you would never get an opportunity to visit your folks ***ku-ma-****yard [at the yards]*— those therapeutically leafy dwellings north of Harare (then Salisbury), where our grandfathers, fathers, grandmothers and mothers labored as gardeners and maids — and part of the reason was because of these bloodhounds. (Newspaper reportage 12)

(17) <$D>: Mm <#> Yeah okay <#> Uh *ndati [I said]* I just want to ask a question-***ka***. (Public dialogue 29)

6.2.3. English verbs with Shona prefixes and suffixes

In the ZimE corpus the frequency of English verbs that were inflected with Shona prefixes and suffixes is 10 times per 100,000 words. The affixes are attached to the roots of words, a process referred to as intraword CS which involves a systematic combination of morphological elements of two languages in a single word (Stefanich, et al., 2019). They include ***ndi-nga-zo-ku****-contact-**a-i***, ***kuti****-explain-**e-ra***, ***zvi-no-mu****-affect-**a***, *appreciate-**ka***,[3] and ***a-ka****-contract-**wa***.[4] Due to the obligatory nature of how mood is indicated on the root and final vowel in Shona verbs, these affixes are common in all constructions (Mashiri, 2002). In the examples above, the final vowels -***a*** and -***e*** are mood indicators.

Examples (18) and (19) illustrate English verbs that were inflected with Shona prefixes and suffixes in English ML phrases.

(18) <$B> <#>But they're fighting among themselves and *tisu tiri* ***ku-***suffer *[we are the ones suffering]* us the residents of Chitungwiza. (Public dialogues 24).

(19) <$A>: Did you have to wait for a few hours ***mu-chi-***monitor *[while you monitor]* or immediately you knew *kuti zvaita [that this has worked]* this has been a success. (Public dialogue 28)

Interestingly, the prevalence of Shona affixes which are classified as late SMs in the 4-M model supports the model's prediction that late SMs are responsible for framing syntactic structure whilst content morphemes play a critical role as primary units in the structure of constituents. Clause coherence was achieved using affixes that were attached to verbs, which are content morphemes. In example (18), the verb *suffer*, which is a content morpheme, was not compromised by the addition of the Shona prefix ***ku-*** *[to]* as a late system morpheme that assisted with clause coherence. In example (19), the content morpheme *monitor* was affixed with the prefix ***mu-chi-*** *[while you]*, where ***mu-*** is classified as a pronoun and -***chi-*** as the conjunction. These late system morphemes (bridges) were responsible for clause coherence.

Inflection of English verbs with Shona prefixes and suffixes also occurred in Shona ML phrases, as shown in example (20) where the final vowel (-***e***) in the English verb (*operate*) is replaced with the Shona suffix -***a***.

[2] The suffix -***ka*** in *question-**ka**, board-**ka***, and *a concerned voter-**ka*** is used for emphasis.

[3] The suffix -***ka*** in *appreciate-**ka*** indicates the modal verb of possibility for something to be appreciated.

[4] The prefix -***ka*** in *a-**ka**- contract-wa* is used for past tense.

(20) <$B>: *Ndiri kukumbira kubvunza kuti panguva dzavakanga vasati vavaoperat-**a** vana ava vanga vachikwanisa kuyamwiswa here naamai vavo. [I want to ask whether the children were breastfeeding before they were operated on?]*

6.2.4. Shona nouns and verbs borrowed for cultural practices and concepts in the ZimE corpus

Interestingly, Shona loanwords or lexical borrowings were prevalent in the ZimE corpus, with the three most prominent ones being *baba [father]* (10 occurrences per 100,000 words), *sadza [maize meal]* (6 occurrences per 100,000 words), and *mai/ amai [mother]* (6 occurrences per 100,000 words). Although *father* and *mother* are English equivalents, *baba* and *mai,* respectively, are preferred and used as a sign of respect in Shona culture, whilst *sadza* is term for an indigenous food that is regularly used. Even though each was only used once, *pada [children's game]*, *ndari [traditional beer]*, *matemba [small, dried fish]*, *madora [mopane worms]*, *muhacha [fig tree]*, and *Museyamwa* (tribal name) are examples of the uniqueness of their use in Shona culture (see Kadenge, 2012). "*Sangoma*" or "*Sangomas*" and "*nyunga*", the terms used to refer to traditional healers, occurred three times per 100,000 words. Shona words were also used to refer to cultural practices such as *kurova guva,* a ceremony performed to bring the spirit of a deceased relative home so that it can be the guardian of the family. The word *Hwindi* is used to refer to people who use shouting and sometimes coercive methods to advertise destinations for public transportation such as buses and taxis. The term *Mushikashika* denotes small taxis used as public transport in Zimbabwe, whilst *Mashurugwi* are gangs that rob mining communities of their minerals and money.

6.2.5. Loanword adaptations in the ZimE corpus

An analysis of the ZimE corpus also revealed instances of loanword adaptations. Even though borrowing and CS are reported to occur along a continuum (Myers-Scotton, 1997), identifying loanword adaptations was possible since the phonological and morphological features of the borrowed items changes (Sridhar and Sridhar, 1980; Kamwangamalu, 1999). In the ZimE corpus, the morphology and phonology of the English words *bus*, *pleasure*, *mango*, *alright*, *cab*, *dollars*, *pension*, and *bales* have respectively changed to *bhazi*, *purezha*, *mengo*, *horaiti*, *kombi*, *dhorazi*, *penjeni*, and *mabhero*, thereby adapting to the morphology and phonology of Shona. Ultimately entrenchment and conventionalization of the words into the Shona lexicon is evident.

6.3. Frequency of functions of code-switching

By analyzing the contexts in which CS occurred, it was possible to determine the frequencies of different types of CS in order to answer research question 6, as shown in Table 4. Only numbers for intersentential CS, intrasentential CS and extrasentential CS were analyzed for functions. Numbers for borrowing were excluded.

Table 4 shows that social functions had the highest frequency in the ZimE corpus (with 37 occurrences). Interestingly, the public dialogue genre in the spoken register had the highest number of social functions of CS. This is because the genre had multiple speakers and, even though the conversations were supposed to be conducted in English, most of the speakers resorted to Shona because they felt comfortable using the language. Even the moderators switched from English to Shona on occasion. It is therefore not surprising that most of the EL islands were recorded in the public

Table 4
Frequency of functions of code-switching in the ZimE corpus for English as the ML (normalized per 100,000 words).

Code-switching functions	Raw frequency	Normalized frequency	Percentage (%)
Switching at a time of vagueness or uncertainty	114	32	24.6
Clarification	15	4	3.2
Emphasis	12	3	2.6
Repetition or reiteration	53	15	11.4
Asking questions	49	14	10.6
Identity marking	45	13	9.7
Code-switching as discourse marking	13	4	2.8
Code-switching to show emotions	26	7	5.6
Social functions (accommodation, solidarity)	132	37	28.5
Unclassified, unfinished sentences	4	1	1
Total	463	130	100

dialogue genre. Switching at a time of vagueness or uncertainty accounted for the second largest frequency in the ZimE corpus, with 32 occurrences. This may be explained by the fact that since English is an L2 in this context, speakers might have experienced difficulties finding English words during conversations, and so switched to Shona. The CS functions of repetition or reiteration, asking questions and identity marking were almost uniformly attested (15-, 14-, and 13 times respectively). Unclassified and unfinished sentences accounted for only one occurrence, which is a very low number. Other functions that occurred less frequently include CS to show emotions (7 times), CS as discourse marking (4 times), clarification (4 times), and emphasis (3 times).

6.4. Functions of code-switching in the ZimE corpus

Several functions were identified in the ZimE corpus with the aim of answering research question 7. The functions for English as the ML are discussed in the subsections below.

6.4.1. Switching at a time of vagueness or uncertainty

An exploration of the corpus revealed that CS was used to clarify. In this case, switching was done at a time of vagueness. In the ZimE corpus, speakers switched to Shona after filled or unfilled pauses. De Klerk (2006) suggests that the hesitation to speak can be a sign that participants had difficulty finding the English words to use during conversations and then resorted to their L1. For instance, in example (21), intersentential CS is used when participant <$L L > starts a conversation in English, uses filled pauses (***eh eh***), and subsequently switches to Shona to explain spending a lot of days travelling to sort out the pension issue. Afterwards, the speaker switches back to English.

(21) <$LL>: Uh uh that's tough. <#>That's a tough one. <#>I think someone else will be frustrated and angry. <#>***Eh eh*** *ndakaita mazuva ndichifambira penjeni yemurume wangu [I spent days travelling to sort out my husband's pension]*. I went from office to office trying to get my <unclear> word</unclear> my husband's benefits. (Private semi-scripted dialogue 38)

6.4.2. Clarification

Switching languages can help clarify a point. Example (22) shows intrasentential CS whereby participant <$A> switches to Shona to describe the type of material that was used to cover the open wound where the conjoined twins were attached.

(22) <$A>: It's a It's a s uh <,,> It's a special type of a a cover that's used not only for conjoined twins but for things like *kana muchi ii muchigadzira vanhu vane ma*hernia *vaya ava munogona kushandisa yatinoti* mesh *iya [when you are fixing hernia patients you can use what we call mesh]*. (Public dialogue 29)

6.4.3. Emphasis

CS is also used to stress or emphasize a point (Poplack, 1980). Intersentential CS is shown in example (23). The first part of the sentence already shows that participant <$Y> was considered to be poor. The speaker further emphasizes the same point by switching to Shona to stress the point of being poor in the next sentence.

(23) <$Y>: They said I, they said **I didn't have anything.** ***Ndanga ndiri murombo*** *[I was poor]*. (Private semi-scripted dialogue 25)

Example (24) also shows emphasis because participant <$RR> first explains that he considers his son to be the *most treasured possession*. The belief that male children are more valuable because they continue the family lineage is rooted in Shona culture. Therefore, participant <$RR> is excited to have a son after having three girls. So, in order to emphasize the importance of the son in his family, participant <$RR> switches to Shona and uses the term "*sarapavana*" *[the one who remains as guardian of the children]* to stress that point.

(24) <$RR>: Uh I will tell you this, my most treasured possession right now is my young boy. I have got three daughters and he is the only boy born this year. <#> He was born this year in a family of *chii [what]* three girls, my children. You see. <#> So I treasure that one. <#> He he is a God given position whom I can say in Shona ***sarapavana*** *[the one who remains as guardian of the children]*. <O> laugh</O> (Private semi-scripted dialogue 44)

6.4.4. Repetition or reiteration

CS can be used to repeat or reiterate a point in order to be as clear as possible. The ZimE corpus shows evidence of Shona words or phrases preceding or succeeding the equivalent words or phrases in English. This is evident in example (25) where participant <$Y> repeats what was said in the first sentence by switching to Shona and saying the same idea in Shona *Um ndinofamba ini [I travel]*, and again repeating the sentence in English. This shows that participant <$Y> is reinforcing the information using extrasentential CS so that the hearer understands what is being said. Example (26) also illustrates repetition where participant <$B> repeats the Shona phrase in English. The repeated parts are shown in bold. Previous research has used the term "hyperclarity" (Williams, 1987) to explain instances where the meaning of a word, phrase or sentence is communicated twice or three times. Earlier studies have also noted that repetition serves a purpose and is a feature of L2 English varieties (see De Klerk and Gough, 2002).

(25) <$Y>: Um I travel a lot. <#> ***Um ndinofamba ini*** *[I travel].* **Um I travel a lot** to Chinhoyi. (Private semi-scripted dialogue 25)

(26) <$B>: <unclear> *Tichiri kupemberera* **We're still celebrating**. </unclear> (Public dialogue 29)

6.4.5. Asking questions

CS serves the purpose of asking questions. Intrasentential CS is evident in example (27) where participant <$O> starts a sentence by using ***Ko*** *[what about]* to enquire about the activities done at church.

(27) <$O>: Alcohol ok. ***Ko*** *[what about]* the activities that you usually do at church? (Private dialogues 8)

In example (28), participant <$C> switches to Shona to show the prejudice faced by women by demonstrating that if a woman speaks her mind, society starts questioning her marital status.

(28) <$C>: The second other problem is<,> beyond trivializing<,> we can even demonize<#> If we hear women speaking<,> we would start to question<,> Netty Musanhi ***akaroorwa here***? *[Is Netty Musanhi married?]* There is <,, > he is speaking like that because she is a divorced woman or whatever woman you want to say<#>. (Public dialogue 7)

6.4.6. Identity marking

CS was also used to show that speakers belong to the same cultural or ethnic group. In this instance, cultural words were used during switches as shown in example (29). To show that they belong to the same cultural group, participant <$P> asks a question regarding culture. In response, participant <$O> switches to Shona to mention the rituals *kurova guva [a ceremony performed to bring the spirit of the deceased relative home so that it can be the guardian of the family]* and *kuroorwa [marriage ceremony]*. In example (29), participant <$O> switches to Shona and expects that the Shona terms used will be understood by participant <$P>. It can be assumed that the two participants belong to the same linguistic and cultural group. In this case, intrasentential CS is used for identity marking.

(29) <$P>: Mm is the new curriculum bringing back our culture?
<$O> <#>Yeah, we are doing heritage, the heritage subjects. <#>We are teach, we are teach, we are taught<O> laugh</O> we are taught about the rituals like ***kurova guva***, ***kuroorwa***, the courtships that were done way back before. <#>Not his online dating and sort of. <#>But we are taught many things of courtship, different things about long ago. (Private dialogues 8)

6.4.7. Code-switching as discourse marking

Interesting observations were made regarding the switches from English to Shona for the purpose of adding discourse markers such as *kana [or]*, *saka [so]*, *asi [but]*, *zvakadaro [still]*, *uye [and]*, and *kunze kwe- [besides]*. Switching to Shona discourse markers occurred 29 times per 100,000 words. Examples (30) and (31) highlight the switches to Shona discourse markers ***saka*** *[so]* and ***asi*** *[but]*, respectively. These two are examples of extrasentential CS because the Shona tags ***saka*** and ***asi*** are added to the sentence but do not violate the grammar of the ML, which is English.

(30) <$RR>: I have got to make sure that I have got to source for water and iron my clothes using electricity at midnight, you see. ***Saka*** *[So]* we are facing a lot of problems. (Private semi-scripted dialogue 44)

(31) <$B>: ***Asi*** *[but]* uh in the interest of serving our listeners I know this the the this has now been raised by two listeners so it is a concern. (Public dialogue 33)

6.4.8. Code-switching to show emotions

There were instances where CS was used to express intense emotions in the ZimE corpus. This is evident in the use of the tag *kikikiki*, a word synonymously used amongst Zimbabwean users to express laughter in writing. In example (32) which represents a written register (newspaper reportage), extrasentential CS is evident because the tag *kikiki* is used to communicate the moods and feelings, by showing laughter (and in this case amusement and disbelief) about the ungratefulness of the camel because it asked the tent owner to leave after being invited to sleep in the tent.

(32) After discovering the duo could not possibly share the available space, the camel cheekily asked its master to sleep outside since he was the smaller of the tent occupants. ***Kikikiki***. What a nerve! (Newspaper reportage 11)

6.4.9. Social functions (accommodation, solidarity)

In some instances, CS was used as a solidarity tool. This was done to show solidarity or to accommodate other speakers. For example, in the public dialogue genre in the spoken register, there were instances where intersentential CS was used to accommodate other speakers who did not want to speak in English. This is illustrated in example (33) where speaker <$B> greets speaker <$F> in English but speaker <$F> responds in Shona and goes on to highlight the preference for Shona during the conversation. In the end, speaker <$B> accommodates speaker <$F> and allows the participant to use Shona. Although participant <$F> knows about the requirement to use English, he asks for permission to speak in Shona, which is granted. Maybe participant <$F> feels comfortable discussing the issue in Shona and even goes on to urge the presenter to use indigenous languages for the benefit of the elderly who might not understand English.

(33) <$B>: My brother how are you?
<$F>: ***Ndaswera maswera here? Totaura neChivanhu ndini mudhara*** **Raba *pa*Norton *apa***. [*I have had a good day, how about you? We will use Shona, I am old man Raba in Norton*].
<$B>: ***Aiwa taurai neChivanhu tinzwe baba***. [*You can speak in Shona dad*]. (Public dialogue 32)

7. DISCUSSION

There are several interpretations for the CS trends involving English as the ML and Shona as the EL observed in the ZimE corpus. The analysis revealed that intrasentential CS registered the highest overall frequency in the ZimE corpus. The fact that intrasentential CS was the second highest type of CS is due to the presence of Shona EL islands in the ZimE corpus, especially in the public dialogues. The high frequency of English words with Shona prefixes and suffixes (28 per 100,000 words) could also be the motivation. In addition, plural morpheme doubling can be considered as a reason for the high frequency of intrasentential CS. This feature has been reported in other L2 English varieties such as Xhosa English (De Klerk, 2006) and Ghana English (Dako, 2002). Intersentential CS was the second highest type of CS in the ZimE corpus, especially in the public dialogues. The nature of the public dialogues allowed for multiple speakers to participate in the conversations. In most instances, the moderator had to switch to Shona in order to accommodate speakers who had responded in Shona.

The analysis showed different functions linked to various CS types. For instance, grammatical and plural marking occurred more at word level compared to above word level. Intrasentential and intersentential CS were mainly associated with social functions, whilst extrasentential CS was used to show emotions and for discourse marking. In other instances, the speakers contributing to the discussions would request to use Shona even though the conversations were conducted on a radio channel that was supposed to broadcast in English. The fact that extrasentential CS was the type of switching used the least can be linked to the fact that, although English is the ML in the conversations, it is used as an L2 by the speakers. Since tag switches have to be inserted at specific points in a sentence, speakers may avoid using extrasentential CS so as not to disrupt the flow of communication. Interesting observations were made regarding the presence of Shona borrowed words in the ZimE corpus. These were mainly words for indigenous foods, cultural practices and traditional games. These words were regarded as lexical borrowings because they filled lexical gaps, as was observed in previous studies (e.g. Sridhar and Sridhar, 1980; Kamwangamalu, 1989; De Klerk, 2006; Bullock and Toribio, 2009).

This study supports the 4-M model's suggestion that content morphemes and early SMs are critical components in sentences. This was shown by the prevalence of plural morpheme doubling during CS in the ZimE corpus. Even though CS occurred in the form of the Shona plural marker (***ma***-), the English plural marker (-***s***) was added to nouns in order to conform to the English ML syntactic structure. Both the CMs and the early SMs played critical roles as primary units in the structure of constituents. In addition, with verbs being classified as CMs, the attachment of Shona affixes to English verbs did not compromise clause coherence during CS.

The incorporation of bound morphemes that occur as below word level switches, single words and multi-word switches that are used in different places which are not adjacent to each other can be cited as the possible reasons for the higher frequency of intrasentential CS in the spoken registers. In general, the spoken registers had more instances of CS compared to the written registers. The spontaneous and sometimes informal nature of spoken registers may make it flexible for speakers to incorporate CS, whilst written language tends to be mostly formal (such as newspaper reportage, editorials, business letters, and academic writing). Here, the functions of CS come into play because speakers may use CS for the expression of emotions, clarification or to emphasize a point.

The lower frequency of switching in the written registers may be attributed to the formal nature of some written registers. For example, out of the 28 switches recorded in the written registers, editorials had two switches which included one borrowed word (*Mashurugwi*), a term used to refer to Zimbabwean gangs that go around mining communities robbing people. Therefore, the use of *Mashurigwi* and other borrowed words in the written registers could be for the purpose of filling a lexical gap. Using the words in texts that undergo editing could be a way of making the readers more easily understand the subject matter.

Discussing the switches at micro-level, the high prevalence of multi-word switches that occurred in the same place (EL islands) can be linked to their higher occurrences in public dialogues. These EL islands occurred in instances where speakers' language preferences were taken into consideration during discussions. The multi-word switches that were recorded in different places occupied second place in terms of their frequency. This can be linked to the functions served by the switches. It could be interpreted that speakers switch to Shona at a time of vagueness or uncertainty and to accommodate other speakers.

The analysis also revealed several functions of CS in the ZimE corpus. Social functions that include accommodation and solidarity had the highest frequencies. Such high frequencies for these instances are not surprising since it has been revealed that the majority of CS occurred in the public dialogues where different speakers participated in the discussions. This may have led to CS for accommodation and solidarity. Switching at a time of vagueness or uncertainty was the second highest function of CS in the ZimE corpus because speakers may have experienced word-finding difficulties during conversations and subsequently switched to Shona. Repetition or reiteration was the third most frequent function of CS. It is highly likely that a tendency to make sure that speakers are understood is part of the explanation.

Regarding the identity marking function, the motivation behind its occurrence at medium level is that speakers wanted to show that they belonged to the same cultural or ethnic group by using familiar terms for cultural practices and concepts. This finding is similar to De Klerk's (2006) conclusion regarding CS in Xhosa English. Although the clarification and emphasis functions were less prevalent in the ZimE corpus, they show that ZimE speakers switched to Shona in order to stress a point and to simplify their utterances for easier comprehension between interlocutors.

8. CONCLUSION

The ZimE corpus data reported here exhibits interesting key findings. There were 537 English ML switches translating to 151 occurrences per 100,000 words. Intrasentential CS was the most frequent switching type, followed by

intersentential CS. The third spot was occupied by lexical borrowing, whilst extrasentential CS was fourth. Loanword adaptations were attested the least in the corpus. With regard to switching where Shona was the ML, intrasentential CS had the highest frequency followed by intersentential CS. Extrasentential CS and lexical borrowing were both attested once. A micro-level analysis of intrasentential CS showed that multi-word switches in the same place (islands), which were attested more in public dialogues, occurred more in English ML utterances compared to Shona ML utterances. Multi-word switches that occurred in different places included instances where English words were affixed with Shona bound morphemes.

Corpus analyses of the frequencies of CS types in spoken registers showed that intrasentential CS recorded the highest frequency, due in part to the inclusion of CS that involves bound morphemes as part of the intrasentential switches. Intersentential CS was the second highest CS type, followed by items that were regarded as lexical borrowing. Loanword adaptations were the least frequent, occurring once in the spoken registers. In contrast, lexical borrowing accounted for the highest frequency in written registers, whilst intrasentential CS and loanword adaptations occupied second spot. Overall, there were more instances of CS in the spoken registers compared to the written registers.

Results showed that CS occurring in the corpus was characterized by (i) instances of plural morpheme doubling on nouns, (ii) the use of English nouns with Shona prefixes and suffixes, and (iii) the use of English verbs with Shona prefixes and suffixes. The analysis showed instances of loanword adaptations and cases where Shona nouns and verbs were borrowed for cultural practices and concepts in the ZimE corpus. Analyzing CS instances in the contexts in which they occurred enabled the determination of frequencies of different functions of CS. Social functions had the highest frequency in the ZimE corpus. The CS functions of repetition or reiteration, asking questions and identity marking were almost uniformly attested. Other functions that occurred less frequently include CS to show emotions, CS as discourse marking, clarification, and emphasis.

Limitations of this study, which provide opportunities for future research, include the corpus size. The compiled corpus was not large enough to be representative of all ZimE speakers. The non-probability sampling technique used meant that participants available for sampling through snowball, convenience and purposive sampling were included, which affected participant diversity. Some areas were not sampled due to time and financial constraints. This in turn has implications for regional representativeness. Future research could thus include compiling a bigger and more representative corpus which is inclusive of all regions in Zimbabwe.

This study has implications in several areas such as linguistic theory, multilingual speech technologies, studies of codeswitching, bilingualism, varieties of English, and language contact. The study contributes to ASR in terms of the development of multilingual speech technologies since there is a need to design and develop pronunciation models, acoustic models and language models that can accommodate different languages. Currently, CS is not explicitly recognized in ASR, resulting in distorted speech processing since some key terms, especially borrowed words are not processed efficiently due to challenges that CS speech poses to speech analyzers, language models, speech recognition and synthesizers and chatbots, among other speech processing algorithms (Modipa, Davel and De Wet, 2013).

An analysis of the functions of CS in the ZimE corpus and a determination of their frequencies is important for a better understanding of the linguistic, cognitive, and sociocultural dimensions of bilingualism in Zimbabwe. The identification of the positions of different types of CS in such discourses as was included in this study, and investigating their frequencies can have implications in five different areas of specialization. Firstly, the implications are apparent in understanding language contact and linguistic patterns, with specific reference to the African context, where the study of positioning, frequency, and functions can lead to new understanding of the underlying syntactic and grammatical constraints occurring in established linguistic rules. Secondly, results from this study can lead to a discovery of new sociolinguistic insights into bilingual proficiency and different functions of CS in different contexts. Examples include the differentiated use between two languages in formal and informal contexts, and in different registers.

Thirdly, this study has implications for cognitive processing, which can thus inform psycholinguistic models of bilingualism (Green and Wei, 2014). Fourthly, position and frequency analyses can shed light on language development and shift in areas of language evolution, language acquisition and intervention strategies (Paradis, 2001). And lastly, this corpus-based linguistic research provides empirical evidence which can inform theoretical models underlined by real-world bilingual speech patterns. Results from this study can also be used in comparative studies in examining universality and variability in CS behavior across different sociolinguistic and grammatical environments, especially in the studies of emerging varieties of English (Bullock and Toribio, 2009).

CONFLICT OF INTEREST

The authors state that there is no conflict of interest.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Faith Chiedza Chapwanya: Writing – review & editing, Writing – original draft, Validation, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Joanine Hester Nel:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Data availability

Data will be made available on request.

ACKNOWLEDGEMENTS

This project was supported by funding from the National Research Foundation (NRF) of South Africa grant number: PSTD240426216115 and the University of Pretoria. Opinions, findings and conclusions or recommendations expressed in this publication are that of the authors. The NRF accepts no liability whatsoever in this regard.

References

Aguirre, M., García-Sardiña, L., Serras, M., Méndez, A., López, J., 2022. BaSCo: an annotated basque-spanish code-switching corpus for natural language understanding. In: Proceedings of the Thirteenth Language Resources and Evaluation Conference, pp. 3158–3163.

Barasa, S., 2016. Spoken code-switching in written form? Manifestation of code-switching in computer mediated communication. J. Lang. Contact. 9 (1), 49–70.

Blom, J.P., Gumperz, J., 1972. Social meaning in linguistic structure: Code-switching in Norway. In: Gumperz, J., Hymes, D. (Eds.), Directions in Sociolinguistics. Holt, Rinehart and Winston, New York.

Bullock, B.E., Toribio, A.J., 2009. Themes in the study of code-switching. In: Bullock, B.E., Toribio, A.J. (Eds.), The Cambridge Handbook of Linguistic Code-Switching. Cambridge University Press, Cambridge.

Chapwanya, F.C., 2022. Investigating the use of articles, modal verbs and selected discourse markers in Zimbabwean English: A corpus-based analysis using the dynamic model. [Unpublished PhD thesis], University of Pretoria.

Dako, K., 2002. Code-switching and lexical borrowing: which is what in Ghanaian English? Eng. Today. 18 (3), 48–54. https://doi.org/10.1017/S0266078402003073.

Déchaine, R.M., Girard, R., Mudzingwa, C., Wiltschko, M., 2014. The internal syntax of Shona class prefixes. Lang. Sci. 43, 18–46. https://doi.org/10.1016/j.langsci.2013.10.008.

De Klerk, V., 2006. Codeswitching, borrowing and mixing in a corpus of Xhosa English. Int. J. Biling. Educ. Biling. 9 (5), 597–614. https://doi.org/10.2167/beb382.0.

De Klerk, V., Gough, D., 2002. Black South African English. In: Mesthrie, R. (Ed.), Language in South Africa. Cambridge University Press, Cambridge, pp. 356–378.

Doke, C.M., 1931. Report on the unification of the Shona dialects. Stephen Austin and sons, Limited, Hertford.

Dörnyei, Z., 2010. Questionnaires in second language research: Construction, administration and processing. Taylor & Francis, New York, NY.

Finlayson, R., Slabbert, S., 1997. "I'll meet you halfway with language"-Codeswitching functions within an urban context. Southern African J. Linguist. 15 (4), 123–134. https://doi.org/10.1080/10118063.1997.9724121.

Fitzmaurice S., 2020. White Zimbabwean English. In: Kortmann, B., Lunkenheimer, K., Ehret, K. (Eds.), The Electronic World Atlas of Varieties of English. (Available online at http://ewave-atlas.org/languages/48. (Accessed 3 October 2024).

Fortune, G., 1984. Shona Grammatical Constructions, Vol. 1. Mercury Press, Harare, 3rd Ed..

Gimode, J., Barnes, L., 2015. A socio-pragmatic analysis of code-switching in the Logoli speech community of Kangemi. Lang. Matters 46 (2), 249–274. https://doi.org/10.1080/10228195.2015.1047893.

Gotosa, K., Rwodzi, M., Mhlanga, G., 2013. Language in education: a critical review of current proposals for official mother tongue use in Zimbabwean classrooms. Int. J. Humanit. Soc. Sci. 3 (14), 88–94.

Grant, A., 2015. Lexical borrowing. In: Taylor, J.R. (Ed.), The Oxford Handbook of the Word. Oxford University Press, Oxford, pp. 431–444.

Green, D.W., Wei, L., 2014. A control process model of code-switching. Lang., Cognit. Neurosci. 29 (4), 499–511. https://doi.org/10.1080/23273798.2014.882515.

Gries, S.Th., 2006. Introduction. In: Stefan, Th., Gries, S.Th., Stefanowitsch, A. (Eds.), Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis (Vol. 172). Mouton de Gruyter, Berlin and New York.

Grosjean, F., 1995. A psycholinguistic approach to code-switching: the recognition of guest words by bilinguals. In: Milroy, L., Muysken, P. (Eds.), One Speaker, Two Languages: Cross-Disciplinary Perspectives on Code-Switching. Cambridge University Press, Cambridge, pp. 259–275.
Gumperz, J.J., 1982. Discourse strategies: Studies in interactional sociolinguistics 1. Cambridge University Press, Cambridge.
Guthrie, M., 1948. Gender, number and person in Bantu languages. Bull. Sch. Orient. Afr. Stud. 12 (3–4), 847–856 https://www.jstor.org/stable/608742.
Haspelmath, M., Tadmor, U., 2009. The loanword typology project and the world loanword database. In: Haspelmath, M., Tadmor, U. (Eds.), Loanwords in the world's languages: A comparative handbook, pp.1-34.
Kachru, B.B., 1992. Models for Non-Native Englishes. In: Kachru, B.B. (Ed.), The Other Tongue: English across Cultures. University of Illinois Press, Urbana and Chicago, pp. 48–74.
Kadenge, M., 2010. Zimbabwean English: a sociophonological exploration. South African Journal of African Languages. 30 (1), 35–51. https://doi.org/10.1080/02572117.2010.10587334.
Kadenge, M., 2012. The indigenisation of English in Chenjerai Hove's novel Bones and Ancestors: a case of lexical and semantic features of chiShona-English. South African Journal of African Language. 32 (1), 11–16.
Kadenge, M., Mugari, V., 2015. The current politics of African languages in Zimbabwe. Per Linguam: a Journal of Language Learning= Per Linguam Tydskrif Vir Taalaanleer. 31 (2), 21–34. https://doi.org/10.5785/31-2-658.
Kamwangamalu, N.M., 1989. Code-mixing across languages: Structure, functions, and constraints. University of Illinois at Urbana-Champaign, [Unpublished PhD Thesis].
Kamwangamalu, N.M., 1999. The state of code-switching research at the dawn of the new millenium: Focus on the global context. South African J. Linguist. 17 (4), 256–277. https://doi.org/10.1080/10118063.1999.9724543.
Koban, D., 2013. Intra-sentential and inter-sentential code-switching in Turkish-English bilinguals in New York City, US. Procedia-Social Behavioral Sci. 70, 1174–1179. https://doi.org/10.1016/j.sbspro.2013.01.173.
Leedy, P.D., Ormrod, J.E., Johnson, L.R., 2021. Practical Research: Planning and Design. Pearson Education Limited, Essex.
Marungudzi, T., 2014. Deploying a sense of plausibility in language choice: the role of English-Shona code switching in Zimbabwean classrooms. Int. J. Eng. Ed. 3 (2), 462–477.
Marungudzi, T. 2016. Some morphosyntactic features and variety status of Zimbabwean English: A corpus-based study [Unpublished PhD thesis], North-West University.
Mashiri, P., 2002. Shona-English code-mixing in the speech of students at the University of Zimbabwe. Southern African Linguist. Appl. Lang. Stud. 20 (4), 245–261. https://doi.org/10.2989/16073610209486314.
Mkanganwi, K.G., 2002. Shona (derivational morphology): an observation in search of a theory. Zambezia Xxix (ii), 174–190.
Modipa, T.I., Davel, M.H., De Wet, F., 2013. Implications of Sepedi/English code switching for ASR systems. In: Pattern Recognition Association of South Africa, pp. 112–117.
Mukenge, C., Chimbarange, A., 2012. A discourse analysis of the use of code switching in the film Yellow Card. Int. J. Linguist. 4 (4), 581–589.
Muysken, P., 2000. Bilingual speech: a typology of code-mixing. Cambridge University Press, Cambridge.
Myers-Scotton, C., 1993a. Social motivations for code-switching: evidence from Africa. Clarendon Press, Oxford.
Myers-Scotton, C., 1993b. Duelling languages: Grammatical structures in code-switching. Oxford University Press, New York.
Myers-Scotton, C., 1997. Duelling Languages: Grammatical Structure in Codeswitching. Oxford University Press, Oxford.
Myers-Scotton, C., 2002. Contact Linguistics: Bilingual Encounters and Grammatical Outcomes. Oxford University Press, Oxford.
Myers-Scotton, C., 2005. Embedded Language elements in Acholi/English codeswitching: What's going on?. Language Matters: Studies in the Languages of Southern Africa 36 (1) 3–18.
Myers-Scotton, C., Jake, J.L., 2001. Explaining aspects of code-switching and their implications. In: Nicol, J.L. (Ed.), One Mind, Two Languages: Bilingual Language Processing. Blackwell Publishers, Oxford.
Myers-Scotton, C., Jake, J., 2009. A universal model of code-switching and bilingual language processing and production. Cambridge University Press, Cambridge.
Nel, J.H., 2012. Grammatical and socio-pragmatic aspects of conversational code switching Afrikaans-English bilingual children. [Unpublished MA dissertation], Stellenbosch University.
Nelson, G., 2002a. International Corpus of English: Markup Manual for: Spoken texts. Retrieved from ice-corpora. net/ICE/spoken.doc. (Accessed: 29 September 2024.
Nelson, G., 2002b. International Corpus of English: Markup Manual for: Written texts. Retrieved from ice-corpora. net/ICE/spoken.doc. (Accessed 18 February 2025).
Ngara, E., 1982. Bilingualism, language contact and planning: proposals for language use and language teaching in Zimbabwe. Mambo Press, Gwelo.
Nyamayedenga, M.S., 2022. Multilingualism as a classroom resource for communicative language teaching: a case of a primary school in Warren Park, Zimbabwe. East African Journal of Education and Social Sciences (EAJESS). 3 (2), 91–101.
Paradis, J., 2001. Do bilingual two-year-olds have separate phonological systems? Int. J. Biling. 5 (1), 19–38. https://doi.org/10.1177/13670069010050010201.
Poplack, S., 1980. Sometimes I'll start a sentence in Spanish y termino en espanol: Toward a typology of code-switching. Linguistics 18, 581–618. https://doi.org/10.1515/ling.1980.18.7-8.581.
Rose, S., Van Dulm, O., 2006. Functions of code switching in multilingual classrooms. Per Linguam: A Journal of Language Learning= per Linguam: Tydskrif Vir Taalaanleer. 22 (2), 1–13. https://doi.org/10.5785/22-2-63.

Schneider, E.W., 2003. The dynamics of new Englishes: from identity construction to dialect birth. Language 79 (2), 233–281. https://doi.org/10.1353/lan.2003.0136.
Schneider, E.W., 2007. Postcolonial English: Varieties around the World. Cambridge University Press, Cambridge.
Sketch Engine https://www.sketchengine.eu/ [software] (accessed 14 February 2025).
Sridhar, S.N., Sridhar, K.K., 1980. The Syntax and Psycholinguistics of bilingual codemixing. Stud. Linguist. Sci. 10 (1), 203–215. https://doi.org/10.1037/h0081105.
Stefanich, S., Cabrelli, J., Hilderman, D., Archibald, J., 2019. The morphophonology of intraword codeswitching: Representation and processing. Front. Commun. 4, 1–27.
Viriri, E., Viriri, M., 2013. The prevalence of codeswitching in secondary schools where English is the official medium of instruction: a case study of Buhera South district. Int. J. Acad. Res. Progr. Educat. Develop. 2 (1), 227–234.
Wardhaugh, R., 1992. An introduction to Sociolinguistics. Blackwell Publishers, Oxford.
Williams, J., 1987. Non-native varieties of English: a special case of language acquisition. English World-Wide. 8, 161–199. https://doi.org/10.1075/eww.8.2.02wil.