A Sociolinguistic Analysis of Variation in a Rural African Community

Chasu in Same District, Tanzania

Rafiki Yohana

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

In the Department of English Language and Literature

(Linguistics)

UNIVERSITY OF CAPE TOWN

DECEMBER, 2009

DECLARATION

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

2. I have used the LSA convention for citation and referencing. Each contribution to, and quotation in, this THESIS from the work(s) of other people has been attributed, and has been cited and referenced.

3. This THESIS is my own work.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signature: _____

Date: 23/12/2009

ACKNOWLEDGEMENTS

It is a pleasure to have so many people to thank for their contribution in producing this work. First and foremost, I am grateful to my creator and sustainer, God Almighty.

I am particularly grateful to my main supervisor, Professor Rajend Mesthrie, for accepting me to be one of his students. He has been tremendously supportive, a role model and mentor. I have been enormously enriched by his insights, critical reading and teachings. I am indebted to him for his moral support and intellectual contribution, but foremost for his patience in showing me the ropes for being a professional linguist. Professor Daniel Mkude, my co-supervisor based at the University of Dar es Salaam, has also encouraged me with supportive academic advice throughout this project. *Havachenisha vose.*

I would like to express my sincere gratitude to the USHEPiA office for accepting me to be a member of USHEPiA family. I have benefited from financial support and it is through a USHEPiA Fellowship that I have managed to do this work. It would be an injustice not to mention Nan Warner, the director of the USHEPiA program for her positive support. *Asante sana mama Afrika*. I also appreciate their consideration in arranging All Africa House to be my home; it has been a quiet and peaceful area conducive to academic enhancement.

I gratefully acknowledge the support of Professor Thomas Hoffman of the University of Regensburg, for assisting me with my data in the Rbrul program and drawing the CART tree model for analysis.

Carrying out this research would not have been successful without the help of the participants themselves. I am particularly grateful to my informants from Mamba Myamba, for lending their valuable time and voices to the data collection procedure. *Mfumwa namuinke lute lwedi vose.*

I do not have enough words to express my gratitude to my parents, Yohana Sebonde and Nazihirwa Lightness for their wealth of love, daily prayers, moral support and the encouragement they have given throughout this project. May the good Lord add on you more days of life. Many thanks to my brothers Hoyange Robert, Sebonde Heriel, Juma Reward and my sister Mchikirwa Sirinael for their continuous encouragement. God bless you tremendously.

Finally but not least, I would like to acknowledge the contribution of Dr.Sarah Rowan for editing this work. I also appreciate the assistance of Ms Nuroo Ismail from the Knowledge Commons, UCT Libraries and Ms. Moonde Kabinga for formatting this project.

Thank you all.

ABSTRACT

This study mainly investigates whether language variation due to sociolinguistic stratification in Western urban speech communities is similar to that in rural African communities, using as a case study the multilingual Chasu of Same district in Kilimanjaro Tanzania. Primarily, the study addresses the question of language use and variation in a multilingual context in which an analysis of the frequency of occurrence of lexical borrowings and code-switching from Swahili and English is undertaken. The study firstly investigates whether the key sociolinguistic variables of social class, gender, style, age and educational levels have as much bearing in explaining the occurrence of code-switching and lexical borrowings in multilingual Chasu. Secondly, the study examines whether social stratification correlates with the phonological variables (s) and (z) in Chasu, along lines established in Western variationist sociolinguistics.

In order to obtain a valuable representative sample of data, the "Labovian" model of the sociolinguistic interview incorporating narratives of personal experience was used. Other complementary techniques such as participatory observation and rapid surveys with wordlists and questionnaires are employed as well. In the context of language contact, the analysis demonstrates that highly educated, young and middle-class speakers are the ones who borrow words and code-switch from word to sentence levels from Swahili and sometimes from English. Through VARBRUL and Rbrul analysis of phonological variation this study reveals further that, while in Western urban communities social factors particularly social class - have significant impact on language variation and change, in Chasu society internal structural factors are the ones that are more influential. Education attainment is a prime external factor in regulating the use of standard variants [z] and [s] against non-standard variants [ð] and [θ] respectively. However, such external social factors are significant only when associated with syllable position, vowels following the variables or the status of the lexical item-i.e. whether a word is borrowed from Swahili or native Chasu words.

List of abbreviations

AdvP	Adverbial Phrase
APPL	Applicative form
CAUS	Causative form
COP	Copula
DIM	Diminutive form
ECL	Ethnic Community Language
EL	Embedded Language
EM	Eastern Maroon
ESR	Education and Self Reliance
FIN	Final vowel
IMP	Imperative form
INF	Infinitive form
MC	Main Clause
ML	Matrix Language
MLF	Matrix Language Framework
NC	Noun Class
NEG	Negation
NP	Noun Phrase
OP	Object prefix/pronoun
PL	Plural form
PP	Prepositional Phrase

PRES/CONT	Present continuous tense
PRES	Present tense
PST	Past tense
PSV	Passive form
RECP	Reciprocity
RLP	Relative prefix
SC	Subordinate Clause
SG	Singular form
SP	Subject prefix/pronoun
SSR	Socialism and Self Reliance
SUBJ	Subjunctive form
T/A PROG	Tense /Aspect Progressive form
T/A	Tense and Aspect form
UPE	Universal Primary Education
VP	Verb Phrase

TABLE OF CONTENTS

LIST OF MAPS

LIST OF FIGURES

LIST OF GRAPHS

LIST OF TABLES

CHAPTER ONE
BACKGROUND TO THE SOCIOLINGUISTIC STUDY OF CHASU

1.0 Introduction

This study takes the standard sociolinguistic view that language cannot be separated from its social context - i.e. from the people who use it, live with it and live in it. Language is not an external entity to human society. Consequently, sociolinguistic study focuses on language variables in relation to human society. It is an essential social phenomenon that exists in and through a speech community (Coulmas 1989). Unlike structural linguistics, which tends to study language as a system that is slightly removed from the users and uses, this discipline investigates and considers language as a functional social entity. It stresses that humans use language to articulate their cultural identity and solidarity, communicate their experiences, ideas and feelings, and still be able to classify and express their social relationships within diverse patterns. Generally there is an interest in the assorted ways in which language and society symbiotically function.

Sociolinguistics rests on the key premise that language is variable and changing in different socio-contexts. Patterns of social stratification such as speakers' gender, age, ethnicity and social class, which compose the socio-cultural, socio-economic interests and other common values of society members, tend to have an impact upon choice of linguistic variables such as particular sounds, grammatical forms, intonation features, words etc. Likewise, speakers' characteristics may influence choices within language varieties pertaining to dialects, registers, code-switching etc. Thus, whilst language is essentially involved in communicating content, in subtle ways it simultaneously helps people to define their relationships to each other and identify themselves as part of one social group or another. This implies that while a speaker's immediate linguistic choices are meant to communicate information, they can at the same time convey important

extralinguistic information concerning the speaker's sex/gender, age, socio-economic class and even sometimes her/his locality, community or nationality, depending on the interlocutor's familiarity with the variety (Tagliamonte 2006:6-7).

Patterns of sociolinguistic variation tend to differ from one speech community to another. Whilst variation theory in monolingual society and the study of code-switching in bilingual communities are thriving areas of enquiry within the discipline (respectively Labov 1972, Myers-Scotton 1993a), they tend to focus on urban environments. The study of sociolinguistic variation in rural based settings of non-Western communities is less prominent in the discipline; this is something which this study aims partly to address.

This chapter serves as a preamble to the thesis by outlining the objectives of the study, giving an overview of the language and the rural African community of the present study. It also describes the status of Chasu in relation to other languages by providing a sociolinguistic profile of Tanzania in which the speech community of the present study is located. It also comments on the implications of the prevailing language policy and language ideologies for the sociolinguistics of the focal community.

1.1 Objectives of the study

The study at hand provides a description of key aspects of sociolinguistic variation in a rural Chasu community. In order to attain this aim, the following objectives form the guideline of this study:

i. To analyse how social variables such as social class, gender, education levels and age are connected to language variation in the context of language contact in the multilingual Asu community of Same District (where Chasu, Swahili and English co-exist).
ii. To investigate whether phonological variation due to social stratification in Western urban speech communities is similar to that in rural African communities, specifically with Chasu of Same District as a case study.

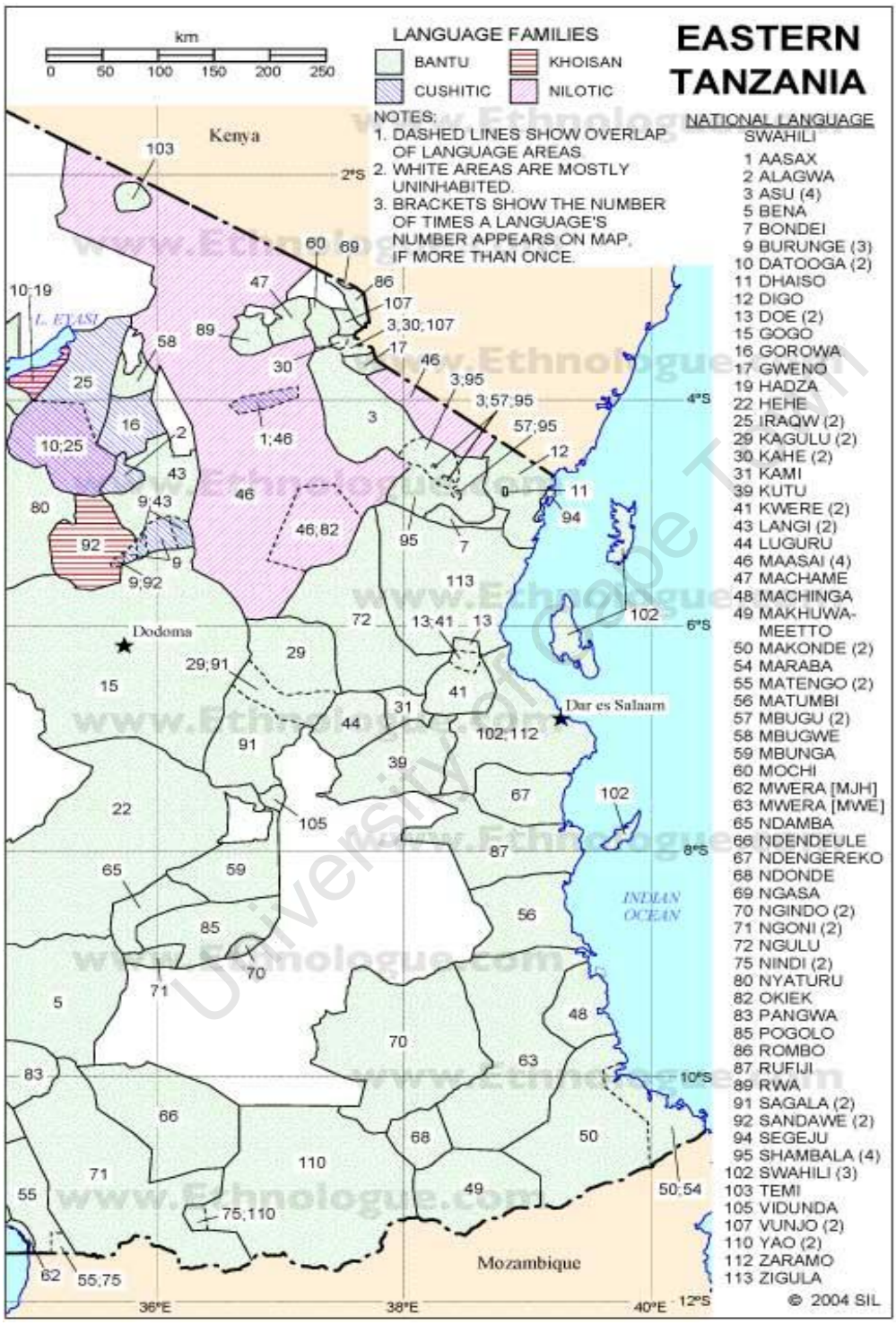
1.2 The community and the language of the study

This section introduces the language of the study, its major dialects and the specific dialect of our interest, the geographical location, and the ethnographic details of the people in that speech community.

1.2.1 The language of the study and its major dialects

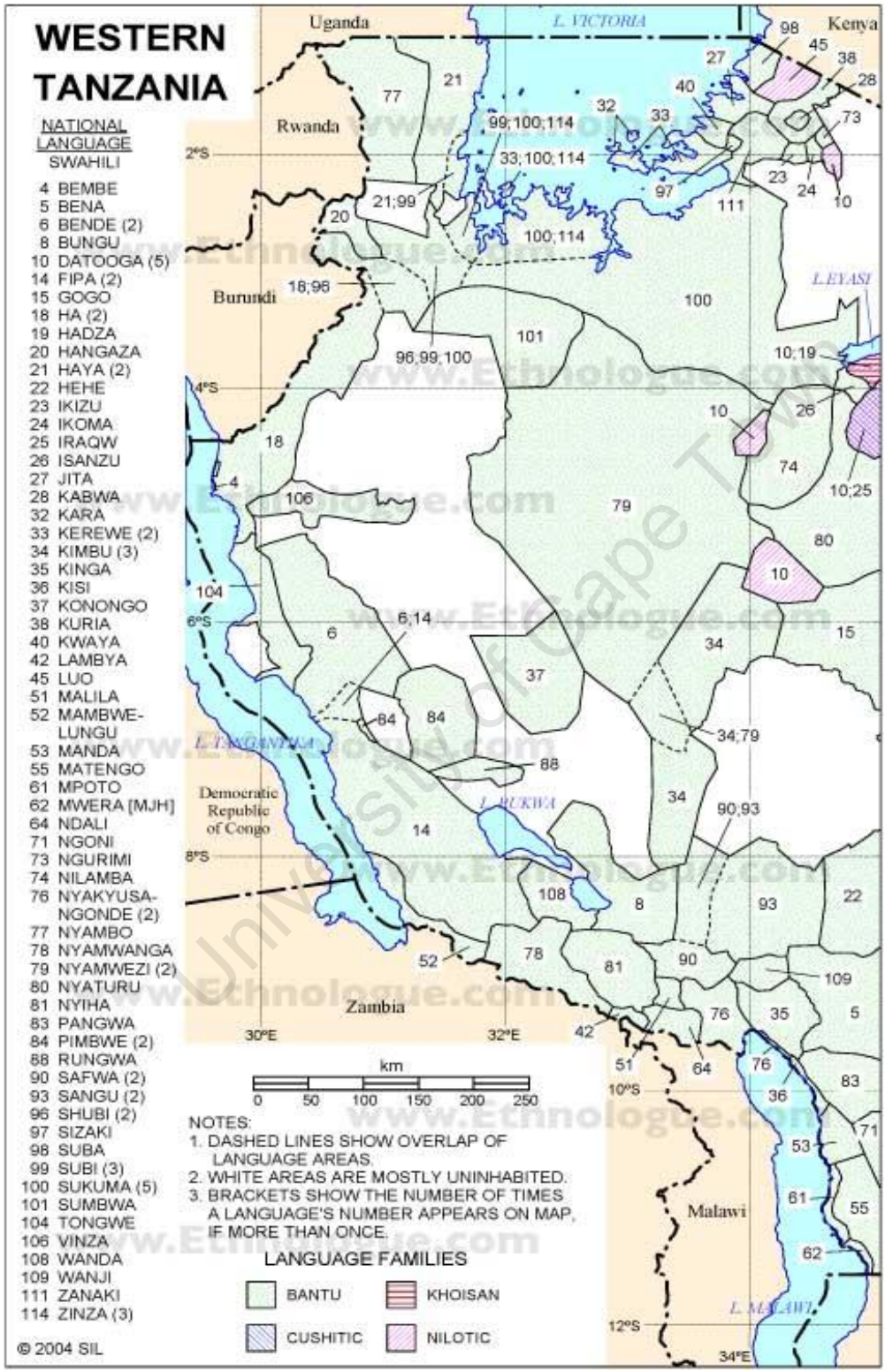
In Tanzania languages fall into four main groups, namely Bantu, Nilotic, Cushitic and Khoisan. As we can see from the Eastern and Western Tanzania maps 1.1 and 1.2 below, the distribution of languages indicates that the Bantu group occupies the greatest part of the country. This is followed by the Nilotic, close to South Western Kenya, comprising languages like Maasai, Luo, Datooga and Okiek. The Khoisan language group comprises Sandawe and Hadza, while the Cushitic group incorporates Asaax and Iraqw, Gorowa, Alagwa and Burunge in Central Tanzania.

Map 1.1: Language Families of Eastern Tanzania



Source: http://www.ethnologue.com

Map 1.2: Language families of Western Tanzania



Source: http://www.ethnologue.com

Our speech community of interest is one of the Bantu groups, the languages spoken widely in East, Central and Southern Africa. The study specifically focuses on the language known as Chasu (see number 3 from the Eastern Tanzania Map 1.1 above), one of the more than 128 Ethnic Community Languages (hereafter ECLs) spoken in Tanzania. It is estimated that a population of almost 500,000 people known as Vaasu or Vampare speaks Chasu[1] in the north eastern part of Tanzania, particularly southern Kilimanjaro. Pare is a name derived from the mountains and hills, which form the greater part of Same District. The inhabitants are called Vampare, after the mountains, or Vaasu, after the name of the language. Of the Chasu community approximately 5% are monolingual, 63% are bilingual in Chasu and Swahili, and 32% trilingual in Chasu, Swahili and English (Lewis 2009). Like any other ECL in Tanzania, Chasu is spoken in homes, several religious contexts, informal settings like funerals, local markets, wedding ceremonies etc. This distribution of language use reflects the sociolinguistic standing and the reputation of ECLs in Tanzania, where it is assumed that individuals are competent from birth with at least one of their ECLs. This language may be supplemented by the acquisition of Swahili or Swahili and English. In the last case all three languages are employed in definite assigned domains (details about language domains are in chapter four).

Several published records of Chasu designate the existence of two main dialects spoken in two areas of Kilimanjaro Region (Mreta 1998:6), namely Mwanga District (Northern Chasu dialect) and Same District (Southern Chasu dialect) where this study is based. Historically, Kimambo (1969) emphasizes that almost sixteen generations ago, groups of immigrants from Taita-hills of Kenya shifted via Lake Jipe searching for new settlements in Ugweno and Mwanga. These groups spread time after time to southern parts of Same District, particularly Bwambo, Gonja, Mbaga, Vudee and later Mamba.

[1] Chasu: Alternate names are Chiasu, Casu, Pare or Athu especially for the Northern dialect. The dialects of Chasu are related to Taveta. Guthrie 1948 classified Chasu from the family of Niger- Congo, Atlantic – Congo, Volta-Congo, Benue –Congo, Bantoid, Southern Narrow Bantu, Central, G, Shambala (G.20) (Gordon 2005).

This is based solely on the evidence of their comparable traditional practices, such as burial customs and worshipping the heads of their ancestors (*kutasa nkoma),* their practices of youth initiations (*ngasu ya mshitu)* and the pottery industry. Another group of speakers of Mbughu (sometimes spelt Mbugu) and the Shambala people from the Usambara Mountains are claimed to have shifted into the South Pare Mountains particularly in Mamba, Hedaru, Makanya, Gonja etc. This is why the Northern Chasu dialect is grammatically affected by northern neighbouring languages such as Gweno and Taita, while the Southern Chasu dialect is greatly affected by the southern neighbouring languages such as Shambala, Mbughu, and Zigua. This is why, based partly on the linguistic features (mainly morphological and phonological) and to some degree geographical considerations, Guthrie (1948) categorised Chasu (the southern dialect which is assumed to be the actual Chasu (Kimambo 1969)) as part of Shambala group in zone G class 22 (Mreta 1994:8). However the origin of the Vaasu either Taita through Northern Kilimanjaro, the Usambara mountains through Southern Kilimanjaro, or the coexistence of Kigweno and Chasu-is still hypothetical.

1.2.2 Salient linguistic aspects of Southern Chasu in relation to its neighbouring Languages

This subsection gives a brief linguistic account of Southern Chasu dialect in relation to its neighbouring languages. It starts by presenting a phonological description of the dialect. Chasu phonemes can be described as follows:

(a) Vowels

There are five Chasu vowel phonemes which can be classified as in the examples below:

[i] Front-high as in *ihandwi* ‘matchet’

[u] Back-high as in *lumi* ‘tongue’

[e] Front mid as in *enga* ‘aunt’

[o] Back mid as in *momo* ‘mouth’

[a] Low as in *amasa* ‘open the mouth’

(b) Consonants

Table 1.1: Chasu consonantal phonemes

	Bilabial	**Labio-dental**	**Dental**	**Alveolar**	**Post alveolar**	**Palatal**	**Velar**	**Glottal**
Plosive implosive	p b			t d			k g	
Pre-nasalized plosive implosive	˜p ˜b			˜t ˜d			ŋk ŋg	
Nasal	m			n		ɲ	ŋ	
Fricative		f v		s z	ʃ ž		γ	h
palatalized fricative		fj vj		sj zj				
Pre-nasalized fricative				˜z				
Affricates					ʧ ʤ			
Pre-nasalized affricates					˜ʧ ˜ʤ			
Trill				r				
lateral				l				
Glide	w					j/y		

Source: Extracted from Kagaya 1989 and Mreta 1998

The Chasu consonants in the above table can be exemplified in words such as the following:

Plosives: <p> [p] *mako**p**a* ‘dry cassava’, <b> [b] *ki**b**ara* ‘guard’, <t> [t] *i**t**ole* ‘mud’, <d> [d] *m**d**o**d**o* ‘aunt’, <k> [k] ***k**isa**k**a* ‘bush’ and <g> [ɡ] *nzi**g**e* ‘locust’.

Nasals: <m> [m] *m**g**uva* ‘sugarcane’, <m> [m] ***m**a**m**a* ‘mother’, <n> [n] *ki**n**aru* ‘water pot’, <ny> [ɲ] *I**ny**awi* ‘cat’, <ng’> [ŋ] ***ng**’ombe* ‘cow’.

Fricatives: <f> [f] *m**f**umw*a ‘traditional king’, <v> [v] **v**a**v**a ‘father’, <s> [s] *mta**s**o* ‘prayer’, <z> [z] *m**z**oro* ‘servant’, <sh> [ʃ] *m**sh**ombe* ‘stew’,< ž> [ž] *ki žo* ‘food’, <gh> [ɣ] *m**gh**oma* ‘female cow’, <h> [h] *i**h**emba* ‘maize/corn’

Affricates: <ch> [ʧ] *m**ch**e* ‘wife’, <j> [ʄ] *i**j**ego* ‘tooth’

Trill: <r> [r] *m**r**angi* ‘throat’

Lateral: <l> [l] *m**l**ao* ‘angel’

Glide: <w> [w] *ma**w**a* ‘liquor’

Pre-nasalized plosives: These are among the pre-nasal sounds of Bantu languages which form a single unit sound. They are formed by a nasal with a stop and they commonly form a homorganic sequence. These are:

Voiceless bilabial nasal stop <mp> [͡ᵐp] ***mp**embe* ‘horns’

Voiced bilabial nasal stop <mb> [͡ᵐb] as in *mra**mb**a* ‘baobab tree’

Voiceless alveolar nasal stop <nt> [͡ⁿt] as in *itu**nt**u* ‘blind person’

Voiced alveolar nasal stop <nd> [͡ⁿd] as in ***nk**inda* ‘banana tree’

Voiceless velar nasal stop <nk> [ŋk] as in ***nk**uku* ‘chicken’

Voiced velar nasal stop <ng> [ŋg] as in ***ng**asu* ‘youth initiation’

Note: all voiceless stops are aspirated when they are pre-nasalized while the voiced stops are not.

Pre- nasalized fricative: There is only one pre-nasalized fricative which is <nz> [nz] ***nz**ao* ‘a male cow’ from [z].

It is imperative to notice that the above mentioned pre-nasalized consonants acquire that status when they occur as the initial syllable of a noun. In Southern Chasu, nouns with the pre-nasalized sounds qualify to be categorized in the 9/10-N/N noun class, which normally takes nouns with nasal sounds as their initial syllable. In Southern Chasu it is rare to find a noun with plosives in the initial word position, as most of them are pre-nasalized to qualify for initial position.

Pre-nasalized affricates: <nch> [ⁿʧ] as in *ich**anch**a* ‘bird’s nest’, <nj> [ⁿɟ] as in ***nj**eta* ‘cob web’. The phoneme [ⁿʧ] occurs rarely in this dialect.

Palatalized fricatives: There are several palatalized fricatives in Southern Chasu including;

<fy> [fj] as in ***if**yo* ‘crowd of animals’, <vy> [vj] as in *ki**vy**ere* ‘in-law’, <sy> [sj] as a *mbo**sy**a* ‘traditional dance’, <zy> [zj] as in *mo**zy**a* ‘in front’.

Palatalization occurs also with a trill <ry> [rj] as in *ki**ry**a* ‘an outcast woman’ as well as in lateral <ly> [lj] as in *ki**al**ya* ‘tomato stew’

Labialized sounds: Sounds of this nature are numerous; generally all simple segmental consonants and pre-nasalized sounds are labialised in this language, except the palatalized segments. It is difficult to labialize other sounds when they are followed by the high back vowel [u].

In his analysis of tense and aspect in Northern Chasu, Mreta (1998) noted the major differences between Northern and Southern Chasu to be at phonological and lexical levels. Earlier Kotz (1909), as cited by Mreta, indicated that at the phonological level, the southern variety has /s/ and /z/ corresponding to the northern variety /θ/ and /ð/. I suggest that in Southern Chasu, these two sounds may be found as a result of the contact with the Northern dialect or with Swahili, where these sounds are used particularly in words borrowed from Arabic. Mreta also observed other differences in phonology in the Northern dialect like bilabial fricative /β/, palatal plosive /ɟ/, velar fricative /ɣ/, and pre-nasalized consonants /mb/, /nd/ and /nð/ corresponding respectively to labio-dental /v/, post alveolar fricative /ž/, velar plosive /g/, and pre-nasalized consonants /mp/, /nt/ and /nz/ in Southern Chasu. Further studies into these dialects may possibly disclose distinctive grammatical properties. The above consonants can be exemplified in words as in table 1.1

Table 1.2: Phonological differences between Southern and Northern Chasu

Southern Chasu	Northern Chasu	English
sakame	**θ**akame	blood
mta**s**o	mta**θ**o	church
m**z**i	m**ð**i	town/city
va**v**a	**β**a **β**a	father
ki **ž**a	ki **ɟ**a	dark
m**g**uva	m**ɣ**uva	sugarcane
mpanga	**mb**anga	cave
ntatu	**nd**atu	three
nzota	**nð**ota	famine

Studies also indicate that Southern Chasu is very similar to Normal Mbugu[2] both in grammar and lexicon, as they are presented as different but closely related languages. Their differences do not go beyond those of dialects. There is some lexical evidence for South Pare affiliation with Normal Mbugu, though sometimes Mbugu is closer to North Pare, as table below demonstrates:

Table 1.3: Lexical and grammatical difference between South Pare, North Pare and Normal Mbugu

South Pare,	Normal Mbugu	North Pare	English
safu	**sa**fu	**θ**ashu	ants
ki**nt**u	kin**t**u	ki**nd**u	thing
nkingo	**nk**ingo	**ng**ingo	skin/hide
Jinka/dindika	dindika	dindika	run
go**nt**a	go**nd**a	go**nd**a	write/carve

[2] According to Mous (2003), Normal Mbugu is a dialect of Mbugu/ Ma'a, a mixed language of Bantu and Cushitic. Mbugu is a hybrid language composed of the Bantu inflectional (prefix and concord) system with Cushitic vocabulary. Derivational morphemes are Bantu and Cushitic. The Bantu influence is from Shambala and Pare/Chasu. Mbugu is sub-divided into two dialects, the Normal Mbugu and Inner Mbugu. The Normal Mbugu dialect is closely related to the Bantu language Chasu in both grammar and lexicon.

Mreta's study also reveals that, unlike Northern Chasu, Southern Chasu has a number of published materials such as a translation of the New Testament (1922-1967), Bible Portions (1910-1960), booklets for readings in Chasu, (i.e. *Masomo a Keri Chasu* (Chasu Second Lessons), *Lesebuch für Evang, Luthschulen in Pare* with popular subjects such as folktales, proverbs, and traditions and customs of the Asu people), the Chasu Translation of the Catechism of Dr Martin Luther (1954), *Kitabu cha Maimbo* in 1928 and *Ngazo ya Mrungu* in 1964. Other materials include a hymn book for the Seventh Day Adventist Church i.e. *Nyimbo ža Mfumwa* and a small grammar book, *Grammatik des Chasu,* written in German by Kotz in 1909.

The southern dialect comprises several sub-varieties based on the distribution of inhabitants, their place of origin, and whether they shifted from the southern or northern regions. These include Kisuji/Kimamba, described by Kotz 1909, and Kimbagha and Kigonja described by Kagaya (1989). Mreta (1998) and Omari (1991) add another variety of Kimakasa-papa, a variety used by Chasu speakers who have lived in isolation since the time of the migration, hence creating their own isolated variety. The present study is focused on the Kisuji/Kimamba variety spoken in Mamba locale, because there has been no linguistic or sociolinguistic study on this variety so far. Nevertheless, the available documentation of this variety in religious books and hymns,, may possibly serve to set a standard of its orthography.

1.2.3 Geographical Location

The Vaasu who are known as mountaineers, live in Pare country, which is said to extend from 60 or 70 miles south eastward from Kilimanjaro along the banks of river Pangani. Same district, whose coordinates are 4°15´S, 37°55´E, borders upon the Usambara Mountains in the Tanga region in its southern ward, and Mwanga district in its northern ward. Hence one may find that the socio-cultural and linguistic features vary within these neighbour groups. Same district is made up of five political administrative divisions, namely Same-Mbaga, Ndungu-Kihurio, Suji-Chome, Hedaru-Makanya and Mamba-Vunta. Same district has a total of 25 administrative wards. The Mamba-Vunta

division is made up of two main sub-divisions, including Mamba with its wards (namely Myamba, Mpinji and Bwambo), while Vunta is comprised of the Vunta and Kirangare wards. The following table indicates the population distribution of each ward in Same District based on the 2003 national census. Myamba ward (in bold: number 17) is our specific area of study.

Table 1.4: Population distribution of Same District

KILIMANJARO : SAME

Population by Sex, Number of Households and Average Household Size

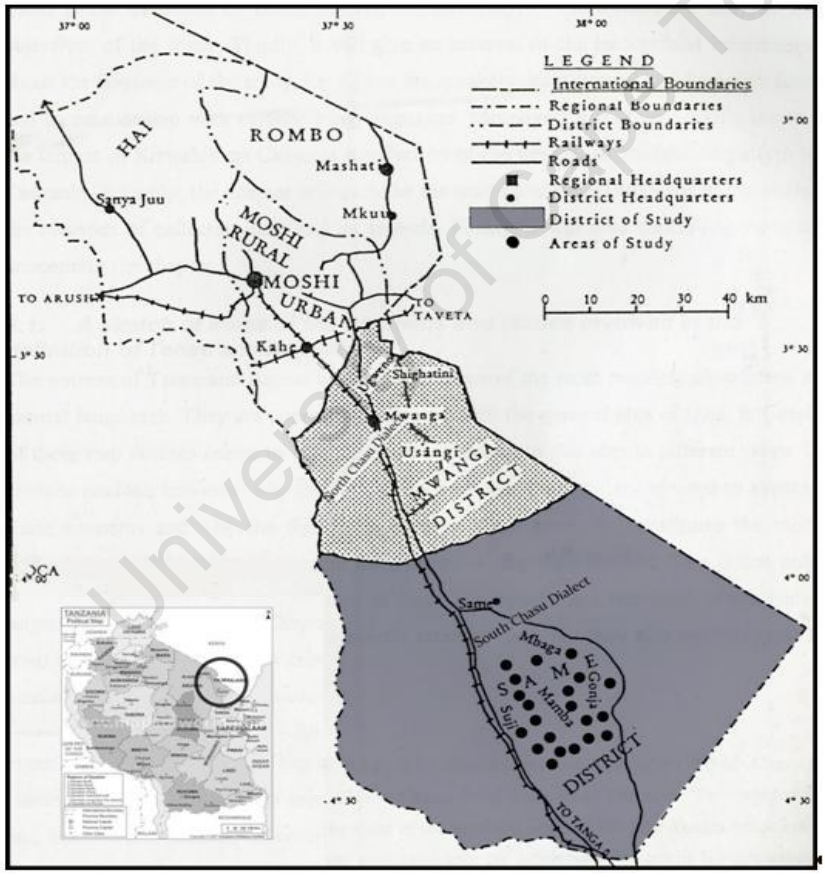
				Population (Number)		Household	
	Ward/Shehia	**Type**	**Male**	**Female**	**Total**	**Number**	**Average Size**
	District Total		**103,520**	**108,805**	**212,325**	**44,474**	**4.8**
1	Same Mjini	Urban	8,168	8,686	16,854	3,957	4.3
2	Ruvu	Rural	4,448	4,052	8,500	2,056	4.1
3	Njoro	Rural	3,995	4,057	8,052	1,564	5.1
4	Kisiwani	Mixed	3,320	3,561	6,881	1,608	4.3
5	Msindo	Rural	2,964	3,110	6,074	1,174	5.2
6	Mshewa	Rural	2,992	3,041	6,033	1,122	5.4
7	Mhezi	Rural	2,202	2,207	4,409	839	5.3
8	Mwembe	Mixed	4,349	4,372	8,721	1,668	5.2
9	Vudee	Rural	2,980	2,969	5,949	1,156	5.1
10	Vuje	Rural	4,628	4,953	9,581	1,773	5.4
11	Bombo	Rural	2,375	2,577	4,952	1,002	4.9
12	Mtii	Rural	5,030	5,272	10,302	2,010	5.1
13	Maore	Mixed	5,470	5,824	11,294	2,726	4.1
14	Ndungu	Mixed	6,841	7,203	14,044	3,251	4.3
15	Kihurio	Urban	3,889	4,200	8,089	1,929	4.2
16	Bendera	Rural	1,851	1,992	3,843	915	4.2
17	**Myamba**	**Rural**	**4,691**	**4,885**	**9,576**	**1,782**	**5.4**
18	Mpinji	Rural	2,991	3,254	6,245	1,243	5.0
19	Bwambo	Rural	4,046	4,403	8,449	1,639	5.2
20	Vunta	Rural	3,661	3,987	7,648	1,505	5.1
21	Chome	Rural	2,518	2,662	5,180	1,032	5.0
22	Suji	Rural	3,974	4,098	8,072	1,419	5.7
23	Makanya	Mixed	4,502	4,676	9,178	1,937	4.7
24	Hedaru	Mixed	9,292	10,242	19,534	4,209	4.6
25	Kirangare	Rural	2,343	2,522	4,865	958	5.1

Source: Government of Tanzania website

Generally, the inhabitants live in small, disjointed villages, often separated by mountains, valleys, forests, small bushes or rivers of various dimensions. Interestingly, while the villages are demarcated by small rivers and hills, in between Mamba and Vunta

there is a large river known as Saseni spouting from the famous water spring/fountain located over the Shengena peak. Shengena peak (2463m) is the highest point of Pare and the Usambara Mountain ridges situated within Chome Forest Reserve, renowned for its colourful flying Butterflies and Birdlife International Projects. It can be reached by bus from the Dar es Salaam–Kilimanjaro highway, crossing from Mkomazi via Ndungu-Kihurio, then up to the mountains. However, during the rainy seasons, one has to learn to walk up these mountains, or rather use a track, as it is difficult and unsafe for the buses to get there, due to heavily muddy and slippery narrow roads constructed through the mountains.

Map 1. 3: Same District



Source: Mreta 1994

We can classify four annual seasons in this countryside: (a) from mid-December to mid-March is the hot dry summer, (b) mid-March to June marks the long heavy rain

season, (c) July to September is the dry cold season and (d) October to mid-December is the season short, but heavy rains.

1.2.4 Ethnography of Asu Community

The information in this section was partly gathered during the pilot research work, via conversations with elderly people. It includes the knowledge of the researcher, who was born and raised in this area. Some of the enquiries from elders were intended as verification of members' assumptions about their community.

The Vaasu form a patrilinear society that was organised into small chiefdoms before the independence of Tanganyika (now Tanzania), when local leadership was abolished by the independent government. The local chiefs were known as *Mfumwa*, and usually came from the Vamjema clan. The Vamjema were known to be rainmakers who were strong and powerful in witchcraft. This situation convinced other clans to give the Vamjema leadership, which was to continue by inheritance. Other clans include the Vabwambo, who are claimed to be the first group to move from Taita to Pare land. Still others are the Vampare, Vankeni, Vamwala, Vamhezi and Vashamba, who are comprised of the Vatewe, Vakirindi, Vambuzii and Vambughu, who moved from the Usambara Mountains. Nowadays it is difficult to identify people according to their clan names except by questioning them. The Vamjema are an exception, as they like to be recognised as official local leaders, and the Vabwambo, who owned the water reserves known as *Ndia*, the sacred forests, (*mshitu wa Ngasu*), where they used to conduct their youth initiations, and the ritual places or shrines known as *itasio* where they worshipped their ancestors.

According to society members, before the coming of missionaries, this community was known for its animistic belief about children. They believed that those children who were born as twins, those whose teeth appeared first from their upper jaws, or the newborn child who turned his eyes to his mother during the birth process were all ill-fated. They were known as *Ndeni Mbivi* "misfortune pregnancy". They were killed by

a specialist in a grinding wooden vessel known as *Kikonti* that was filled with water, before the child was placed in it with his head upside down. Otherwise, they could be taken to what is known as *Ibwe la Vana*, 'children stone', a special stone for killing "misfortune children". They were taken there when they were asleep, left there to wake up, fall into the surrounding marshland and die.

Worshipping procedures differed from one clan to another. The Vamjema worshipped a certain big tree known as *mvumo*, while the Vamtewe clan worshipped a tree known as *mdarya*, but the famous group was the Vabwambo, who worshipped the heads of their ancestors, *kutasa nkoma*. This was the commonly-known way of worshipping and making rituals among the Vaasu. The heads were taken from the grave a few days after burial when they start to decompose; the procedure was accompanied by some rituals. The heads were hidden in their shrines (*matasio*), in the caves or in the roof of the shrine leader. The clan members would gather annually for a ritual in their *itasio*. They believed that it was through these rituals that they could get rid of incurable diseases and misfortunes.

In Chasu society, various initiation ceremonies also differ from one clan to another. However, the most famous ceremony was for the youth initiation known as *Ngasu ya Mshitu* for boys and *Kuaekwa* for girls. The procedures imparted knowledge and skills which could serve the young ones during their parenthood. It was an abomination for a girl to fall pregnant before initiation and marriage, as she would be taken as an outcast known as *kirya.* *Kirya* status was perceived to bring shame to the parents. Hence they were obliged to make a small hole behind their house for the *kirya* to pass through, as it was believed that more misfortune would occur if she passed through the door. She was then taken to survive or die in the forest alone. For boys, *Ngasu ya Mshitu* took place in the forests of the clan. In certain circumstances, cowardly boys could die of fear in the forest as the procedures were accompanied by threatening boys with the sound of lions. The youth initiation procedures went hand-in-hand with circumcision for boys and girls. After the initiation, the young men could be free to look for life partners. A man could marry by giving five or six cows as a dowry, depending on

the nature of the clan. The marriage ceremonies could take place in traditional ways, where *dengelua*, a traditional local brew made of sugarcane juice (*puya*), would be used in celebration. People would perform *mbosya* and *kisesya*, forms of traditional dancing.

Except for the marriage practices, all other traditions were abolished by the missionaries as well by the government after independence. However, the Vabwambo still secretly perform some annual rituals in their shrine, (*itasio*) or when the clan members have problems associated with misfortunes. The government and western religious doctrines associated these practices with paganism and animistic beliefs. However, those forests where the rituals and the initiation took place are reserved as sacred places till today. This has served as a means of environmental conservation. Although the Vamjema chiefdom was abolished, the clan still exists and currently leads the society, largely in the environmental conservation campaign. Nowadays the society has been converted to Christianity and to Islam for a few individuals.

The Chasu community had schools, which were established by the missionaries who arrived in 1903. However, after independence, like many other rural communities in Tanzania, this place was marginalized in terms of development, as it is sometimes inaccessible due to its poor transport, especially during the rainy seasons. Nevertheless, the villages are currently equipped with telecommunications and electricity provisions; this has started to change the lifestyle of the people. There is an interaction with the outsiders who come for business and small scale investments. More educational institutions are being established, with at least one secondary school in each ward. These are well attended. As the local language has thereby come into contact with other languages, and many young people are now at school, traditional and once stable patterns of language use have started to change.

1.3 Social groups

Social class and social stratification are the expressions utilized to refer to any hierarchical ordering of groups within a society (Trudgill 1974a:35). Social class is generally taken to be an aggregate of individuals with similar social and/or economic characteristics and social mobility. Individuals can be classified with respect to the rest of their society by quantifiable characteristics like income, education, occupation, residence or lifestyle and beliefs (Milroy 1980, 1987:99). Other social variables such as race, ethnic background, age and sex may also play important roles in the formation and continuation of social groupings. It is from these social groups that language can be perceived as a means of social interaction, and as a possible mode of detaching individuals from one group to another.

Nevertheless, these characteristics of social groupings can also be ordered in accordance with the way they are utilized in and evaluated by a particular society. In the industrial world, the non-linguistic variables such as lifestyle, attitude and beliefs as well as differential access to power, prestige and wealth can be used to identify social groups. In other parts of the world things might be different. In India, for example, society is stratified into different castes, clearly separated from each other with hereditary membership with rare possible movement from one caste to another, leading to caste dialects (Trudgill 1974a:35-36). This makes it difficult to uniformly classify social groups in different societies. Rural inhabitants may be structured into different social groups compared to urban dwellers. Singler (2009) suggests that African communities, especially in West Africa, are better classified according to ethnicity, community membership (rural in opposition to urban) and gender than by social class.

Taking this into consideration, rural Chasu speakers of the Southern Dialect are likely to exhibit differences in linguistic variations from what we expect of urban socio-economic groups. However, unlike groups in India or the western urban world, the groups studied in rural Mamba can be classified according to daily economic activities associated with the pastoralism, peasantry, owners of small business and professional

activities (teachers, nurses, religious and political leaders). It should be noted that these economic activities are the basic source of earnings for each group. It is through their earnings that one can rank them in a social hierarchy from lowest to the highest.

The group of individuals with professional activities includes nurses, rural assistant medical doctors, teachers (in primary and secondary schools), religious and political leaders. The occupations they hold may be also associated with the level of education they have acquired, which ranges from ordinary secondary school to certificate, diploma, and bachelor degree levels. Although they earn a regular monthly salary, some of them also practise farming and small business. They are perceived by other society members to be of high status. Together with Chasu, this group is exposed to both Swahili and English usage, though to different degrees of acquisition and proficiency.

Another group includes the businessmen and women who own shops and kiosks selling construction materials, clothes and foodstuff. Some own flour milling machines and restaurants, while others own capital, which enables them to conduct business during the harvest season. Despite the fact that most of these are not highly educated, i.e. having only primary school education or in few cases secondary education, they travel to the city centres to purchase commodities and sell crops. They have thus gained some exposure to urban lifestyles. They use Swahili in most of their business transactions on their travels. Their status is considered lower than that of professionals because of their education levels.

Like other rural areas, land ownership for cultivation in the Chasu community is titled to the clan or to the head of the household, which in this patrilinear society is a man, leading most of the society members to practise farming. There are people who depend fully on agriculture as an essential source of livelihood. A good number of the inhabitants in this rural community depend on agro-farming, and are known to be cultivators of sugarcane, ginger and coffee as their cash crops, whilst maize, beans, bananas, yams, sweet potatoes, cassava, millet and various types of vegetables and fruits are cultivated for their daily consumption. Except for the cash crops, which involve

outsiders coming into the area for business transactions, the rest of the crops are for household consumption with rare local exchange. Due to the mountainous nature of the landscape as well as unpredictable rainy seasons, the peasants depend on practising irrigation and the terracing method of cultivation. Together with the Chasu language, this group is exposed to a little Swahili acquired during their primary school education.

Animal keeping is practised by another group of people living in isolated settlements to allow free-range grazing of sheep, goats and cows. This is a small group inhabiting the southern part of Mamba with its savannah terrain, which is full of pastures. The pastoralists rarely practise agro-farming. Instead, live-stocking is their main source of earnings and consumption as well. Possessing a large number of livestock is perceived to be prestigious. However, the owners do not enjoy a high standard of living, as children do not go to school, there are no good houses, and it generally takes time for the pastoralists to sell as their animals to purchase other goods and services. Some of the animal keepers are said to be immigrants from the Mbugu area in the Usambara Mountains, as well as Muhezi in the Mbaga area. It is important to acknowledge that, unlike pastoralism in Masai society in Northern Tanzania, where both husband and wife are grouped as pastoralists, in Chasu animal keeping is only for men, while their wives are housewives or practise farming.

The achievements of the last two groups depend much on the climate of the settlement area. Individuals who are engaged in pastoralism and farming rarely move to the urban areas, due to the demanding nature of their work. They lack the opportunity to contact people who speak Swahili or English, and hence remain Chasu speakers, with a little Swahili spoken in the churches or mosques they might attend. There is also unpractised elementary English learnt in their primary school, if they managed to grasp it at all. Due to the isolation of the pastoralists, their Chasu might remain a unique variety.

Generally, these are social groups, within the Chasu community, and they form a basis for social categorisation of the community for sociolinguistic purposes.

To my knowledge only two serious academic studies to the language have been conducted. The first is *Tense and Aspect in Chasu Verb* (Mreta 1994), which focuses on the Northern Chasu dialect. The second is *A Classified Vocabulary of Pare Language* undertaken by Kagaya (1989) in some varieties of Southern Chasu; Kimbagha and Kigonja. This led to the sub-studies *A Study of the Tonal System of the Gonja Verbs and Nouns-A dialect of the Southern Pare Language* (1989) and *A study on the tonal system of the Mbaga Dialect of Southern Pare Language* (1989). These sub-studies are based on only two informants, one from each variety. No sociolinguistic study has been conducted on the Chasu speech community at large or the Southern Chasu dialect in particular. It is for this reason that this thesis focuses on the Southern dialect of the Chasu language, with special reference to the Same rural community.

1.4 Chasu in relation to other languages in Tanzania

This section presents a synopsis of the basic information on the relationship of Chasu to other languages in Tanzania. The first section discusses the position of Chasu, one of the ECLs, in the sociolinguistic profile of Tanzania, and the second focuses on the status of ECLs in the language policy of the country.

1.4.1 Sociolinguistic profile of Tanzania

The African continent presents the most complex linguistic portrait in the world, owing to the number of languages, the diversity of language families and the functions assigned to various languages in the same country. The language count per country varies from 4 in Rwanda to 516 in Nigeria. Tanzania is said to have more than 128 languages, with its linguistic diversity index of 0.965 placing it first in Africa and fourth in the entire world. Papua New Guinea 0.990 has the highest linguistic diversity in the world, followed by Vanuatu 0.972, then Solomon islands 0.965 (Gordon 2005). (Lieberson 1981), as cited by Alfred and Romos (2007:6), describes the linguistic Diversity Index (DI) as the level of language distribution in relation to the population within a country. DI indicates that when two people of the same country are randomly selected there is a

probability that they speak different mother tongues. If two people have different native languages it indicates total DI represented as 1, which is the highest possible value. The lowest value 0, occurs when everyone speaks the same native language, to indicate the complete absence of diversity. The equation for DI is:

$DI = 1 - \Sigma (P_i)^2$, where:

P_i = the percentage fraction of the total population which is comprised in i^{th} language group.

i = 1 to n, where n is the number of languages that comprise the society.

Σ = the summation of $(P_i)^2$ for all i

A greater variety of languages should increase the value of the index, but as the proportion of language group decreases, its contribution to diversity index should also decrease. In this manner, countries with many language groups of roughly equal size, like Tanzania, will show relatively high linguistic diversity, whereas countries with comparable numbers of languages, but with one or two dominant languages, like the United States of America, will show relatively lower linguistic diversity.

In Tanzania particularly, this complexity is not only reflected in the number of languages spoken by communities of varying sizes, but also by the comparative power, status and functional domain assigned to various languages. Being a multilingual society, language use in Tanzania extends from ECLs through to Swahili, the national and official language, to English, the official and international language. Though cultural policy does not clearly state the functions of ECLs in Tanzania, they are assumed to be prominent at home (or within families), specifically in rural communities, for intra-ethnic and intra-cultural communication and other informal domains of language use (Mochiwa 1979, Ndezi 1979 & Mekacha 1993:9). Swahili is the first language of some coastal people and residents of Zanzibar, as well as the younger generations of the urban inhabitants. It is a second language to more than 30 million rural Tanzanians (Lewis 2009). It is the language, which is prescribed by law for use in most governmental business, in addition to being the medium of instruction in all government primary schools and a subject in secondary and tertiary education.

Although Swahili is a national language of Tanzania, spoken by a large bilingual population, knowledge of the language and the extent of bilingualism are not uniform, especially in the rural communities. Based on the survey done in north east Tanzania in Pare District (now Mwanga district), which involved Northern Chasu, O'Barr (1971:4) identifies three major trends in the knowledge of Swahili in Tanzania:

i. Men almost universally tend to have some knowledge of Swahili compared to women who use the vernacular.
ii. Younger people have greater capacity for using Swahili than older people.
iii. The more literate and educated a person is, the more likely he is to be a fluent speaker of Swahili.

Although this survey is now almost 40 years old, it is still valid. To this one might add an urban-rural dichotomy, as Swahili tends to be spoken much more in urban than in rural areas. Due to inter-ethnic marriages in urban locales, children use Swahili as their first language, as they are unable to learn two different vernaculars from their parents, who also use Swahili as their intermediate language. The language also dichotomizes people according to degree of education, age, gender and coastal versus interior patterns of language use. Generally, this gives a clue to the patterns of language use we can expect in the rural community.

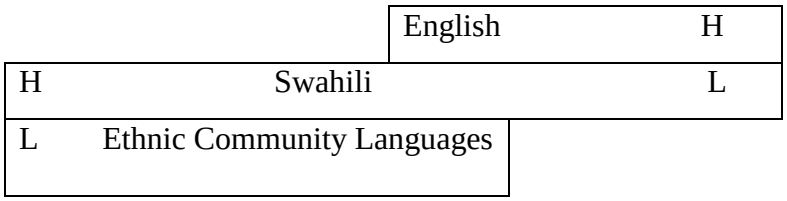
These facts guided some scholars like Fishman (1972) and Abdul-Aziz Mkilifi (1972) to introduce a triglossic model in describing language use in Tanzania. Triglossia is an extension of Ferguson's (1959, 2003) concept of diglossia, to characterise the functional distribution of language in multilingual society. Speaking of language use in Tanzania, Abdul-Aziz Mkilifi (1972:198) claims that;

> a typical triglossia situation would be found where there exists side by side (a) regional or vernacular languages whose role is in oral intra-group communication, (b) a local standardized lingua franca which is used extensively in the education system, mass media and government administration and (c) a world language.

From Tanzania's position, this model proposes that languages are assigned roles in complementary distribution. The ECLs are acquired and used at homes and in informal domains. Swahili is first learnt in primary schools and used by the majority of people. It thus fosters unity and has become a symbol that expresses and mobilizes national pride. It is used in official day-to-day administration to ensure the smooth functioning of the political, social and economic systems of the nation. It also guarantees access to these systems to different social groups and equal opportunity to participate in them at the national level. The English language is for higher education, the high court and the court of appeal, diplomacy, foreign trade and other international business dealings that cut across international levels.

Fasold (1984:44) uses Mkilifi's idea of triglossia to further introduce a "double overlapping" model when he describes the language profile of Tanzania. There is an intersection between two developing diglossia situations, one involving Swahili with some ECLs and the other involving Swahili and English. Swahili is involved in two diglossia systems: as High language with the various ECLs, and as a Low language with English as High. This description is based on the assumption that not all people can access English or ECL, but all Tanzanians can access Swahili. Swahili can be considered as the language of the entire Tanzanian community, while English or ECLs are for specific groups. This describes Tanzania's language profile as a double overlapping diglossia as the figure below indicates.

Figure 1.1Double Overlapping Diglossia in Tanzania



Source: Fasold (1984:45)

Using Abdul-aziz's idea of triglossia or rather Fasold's "Double Overlapping diglossia" in describing Tanzania's language profile, Mekacha (1993) cautions that, we should not assume that the community's verbal repertoire can result in full accessibility to the languages for all community individuals. English is only fully accessed at the higher learning institutions, which are aimed at a minority of citizens. English has little communicative status for the majority of Tanzanians. ECLs are acquired by the rural inhabitants and some individuals who try to maintain their vernaculars in the urban areas. Consequently, to a majority of urban dwellers, ECLs have no communicative status, especially in families where intertribal marriages have taken place. Swahili is perceived to cut across all individuals in the community. However, there are a few individuals who cannot converse in Swahili in rural areas, and are consequently excluded from the domains in which Swahili and English are dominant.

Mekacha (cited by Grimes 1996) suggested that Tanzania had the pattern of a tripartite linguistic repertoire, as it is acknowledged that Swahili is a second language for the overwhelming majority of the Tanzanian population, acquired after or, increasingly, alongside one of the ECLs. For most Tanzanians, bilingualism is a way of life, in so far as they are competent speakers of Swahili as the language of wider, national communication and of an ECL such as Sukuma, Chasu, Haya or Maasai.

Despite these agreements, it is clear that the pattern of language use in Tanzania involves three levels. The ECLs, which could be termed "minority languages" in terms of their status, prestige and literacy, are known by most people particularly in rural

settings. Swahili occupies the second level as the national and one of the official languages, which commands prestige as a standardised language. English is used at the third level for international communication and is highly prestigious in terms of economic and political power.

There are no open conflicts or social clashes regarding the functional distribution of the languages according to their relative power and status. However, individuals realise on a daily basis that by virtue of their political and economic status, English at one level and Swahili at another become a prerequisite for access to power and upward mobility within a society (Rubagumya 1989:109). There is a rapid functional shift from the non-prestigious languages to the prestigious as a result of economic and political pressure within the country and from the international level. Due to a lack of prestigious status the ECLs have been marginalized. This will be discussed in the following section, which deals with the status of ECLs in language policy.

1.4.2 The Language Policy of Tanzania

This section presents an overview of the language policy in Tanzania with regards to the functional distribution of languages and in terms of power and status. It provides the historical and ideological perspectives attached to the language policy and diverse recognitions given to Swahili against the ECLs before and after independence.

(a) Language Policy before Independence

Language use in Tanzania before independence can be explained in terms of the languages that the colonists preferred to promote and use. For the purposes of this study, this will be traced from German colonialism, then British colonialism and finally during the struggle for independence.

During the German colonial period (1885-1918) in Tanganyika, part of their East African territory, Swahili was found to have been spread in many parts of the country due to the ivory and slave trade caravans from the coast to the hinterland, run previously by

the Arabs. When Germans took over the territory they sometimes hesitated to promote Swahili, seeing it as a threat to Christianity as it was connected with the Islamic religion. However plans to introduce German in some schools were not successful (Rubagumya 1990:6). The policy shifted to the promotion of Swahili as a lingua franca, with emphasis on education and administration. For both Germans and Africans, one had to know Swahili in order to be employed as a civil servant or a member of the armed forces. German missionary scholars such as Carl Velten, Krapf and Buttner wrote the grammars of Swahili and collected several manuscripts in Swahili and Arabic scripts of the pre-twentieth century classical literature. This facilitated the rapid spread of Swahili in education (Mohamed 2003). The missionaries also translated some religious texts like bibles and hymn books into the ECLs, particularly where they had new converts.

After the 1st World War, Tanganyika was mandated to the British, who continued to encourage the teaching of Swahili and other languages. Swahili was used as a medium of instruction in primary schools and as a subject in secondary schools. English was introduced as a subject from standard three of primary education (grade three in international terms) and used as a medium of instruction from middle schools (standard five-eight) and in secondary schools (nine-twelve) and in tertiary education. Swahili was well established in local government and local courts, while English was used in central government. Since Swahili contained various dialects, the East African governors decided to standardise the language when they realised that it was feasible as a lingua franca. In January 1930, the Inter-territorial Swahili Language Committee was established to promote standardization. The committee had to select the appropriate dialect for education, standardise the orthography, control publication of books and dictionaries, secure uniformity of grammar, give assistance that may encourage native-speaker authors to translate books into Swahili and ensure the correctness of the Swahili before the publication of those textbooks and other literary materials (Kiango 2005). It is therefore apparent that during colonialism, more effort was made in promoting Swahili compared to English and other ECLs.

During the struggle for independence in the 1950s, the TANU (Tanganyika National Union) political party used Swahili as its working language (Blommaert 1996:13). Members of TANU were taken from different ethnic groups, as it adopted an anti-tribal egalitarian policy. Since Swahili was used to mobilise and unite political party members, it became associated with egalitarian and national ideology. It was established as the language of politics and independence, and was used at all public occasions and in mobilizing events organised for the independence struggle. Swahili stood as the language of all Tanganyikans aiming to restore political and economic power to a classless society of Africans. Using ECLs connoted tribalism and anti-national building. Blommaert **(**1997) claims that the use of English signified imperialism, oppression and neo-colonialism, while Swahili stood for independence, Africanhood, self-confidence and freedom. Generally the functional use of language was highly ideologised.

(b) Post Independence language policy and practice

Shortly after the independence of Tanganyika in 1961, Swahili was declared the national language (Rubanza 1979:33). It is widely held that President Nyerere's speech of 10th December in Kiswahili was a landmark in its development, as Amri Abeid comments, cited in Whiteley (1971:166), that "an unprecedented thing that happened on that day was that his Excellency delivered his speech in Swahili. That moment was truly the beginning of a new era in the history of development and the running of the affairs of the government…."

After this president's inauguration in 1962, Swahili was used in Tanganyika as the language of the Parliament. It was used in public sectors, and directives were given to both governmental and parastatal offices requiring them to use Swahili as the official language (Whiteley 1969). It also became the medium of instruction within the entire primary education system. Swahili then became the national language and an important tool used in the state affairs of the independent government. The TANU party and government leaders continued to mobilize people against poverty, disease and ignorance. Since education was promoted as a major goal in these campaigns, a rapid expansion of

education opportunities to Africans through Universal Primary and Adult Education occurred. Swahili was used in these operations as a means of communication.

One of the most politicised movements that affected language policy was the Arusha Declaration of 1967, which came with the policy of Education and Self Reliance (ESR) that was part and parcel of the philosophy of Socialism and Self Reliance (SSR). Egalitarianism was the central ideology of education and self reliance, and attempts to make education accessible to all community members were viewed as a key to promoting social cohesion. The education that was given was complete in itself, as it anticipated preparing people to work in rural areas. Hence the language of instruction which was most effective for this campaign was Swahili and not English (Morrison 1972). Through education and self-reliance, Swahili was taken as a tool for building and attaining a socialist political ideological state. Speaking other languages like English would be perceived as antisocialist elitist behaviour (Blommaert 1996 & 2005).

Various institutions were established by the government to strengthen Swahili (Kiango 2005). These include the policy making body, The Directorate of Culture and National Language. The National Swahili Council was established in 1969 as a policy-implementation body, and so too was the Institute of Kiswahili Research, formerly known as the Inter-territorial Swahili Language East Africa Committee.

As suggested earlier, after independence ECLs were regarded as a threat to the national unity, since they carried the potential to promote tribalism and ethnic division. It was perceived especially by the government that when struggling to mould the nation it would be undesirable to encourage ethnic languages. This left the ECLs with no viable future in most formal domains and, sometimes, in those domains in which they could have been expected to be used, such as family talks, marriage ceremonies, local markets etc. They were consequently marginalized to informal domains like traditional functions and ceremonies in the villages. It was during 1990s, after multipartism was introduced in Tanzania, that a new, more tolerant official attitude towards ECLs came into being (Legére 2002). The current Cultural Policy, inaugurated in 1997**,** has officially

recognised the existence and value of ethnic languages and has tried to define their roles in the society. Although the functions are not well classified, ECLs are now considered to be used as the resources for the development of Swahili. Individuals are encouraged to use and be proud of their ethnic languages. More encouragement is given for research, writing and translating languages, writing dictionaries and grammar books, publishing and disseminating materials in ECLs. The articles can be read in the Cultural Policy of The United Republic of Tanzania (1997:33) as follows:

1.1.5 Vernacular languages shall continue to be used as resources for the development of Kiswahili

1.2 Vernacular languages

1.2.1 Our people shall continue to use and be proud of their vernacular languages

1.2.2 Communities, private and public organizations shall be encouraged to research, write, preserve and translate vernacular languages into other languages

1.2.3 The writing of vernacular language dictionaries and grammar books shall be encouraged

1.2.4 Public and private organizations shall be encouraged to publish and disseminate vernacular language materials.

Every country faces its own difficulties when implementing a language policy. Tanzania is often upheld as a model of successful language planning in favour of indigenous African languages. It is true that Tanzania has made a lot of progress in cultivating Swahili as a viable national language which is used in almost all spheres of national life and in education. However, this success is usually exaggerated by outsiders

(Rubagumya 1990). Like many other ex-colonial African countries Tanzania considers the language of the former colonial ruler, English as the most suitable medium of instruction above primary education level. English is still regarded as a prerequisite for scientific and technological development as well as for international communication. This makes the sociolinguistic profile more complex, like that of many other African countries.

However, the provision presented above by the Cultural Policy concerning the ECLs grants an opportunity for language research projects. The 'Languages of Tanzania' project (LoT) that was initiated in 2001 is a product of this policy. The project deals with languages spoken in Tanzania besides Swahili. It produces a description of each language and a list of vocabulary, determines its genetic classification and the number of speakers, and places it on a language atlas, showing where it is spoken in the country (Legére 2002). The project has managed to identify the existing languages of Tanzania. The largest 20 languages, are as the list below in table 1.4 indicates, arranged in descending order of numbers of L1 speakers. The entire list of Tanzanian languages is attached as appendix 2.

Table 1.5: A list of the languages of Tanzania by numbers of L1 speakers, 18 July 2007 (adapted from Languages of Tanzania Project (LOT))

SN	Language	Speakers	SN	Languages	Speakers
1	Sukuma	5,189,951	11	Fipa	712,786
2	Swahili	1,379,692	12	Iraqw	602,661
3	Ha	1,228,864	13	Bena	592,370
4	Gogo	1,023,618	14	Sambaa	565,276
5	Nyamwezi	958,898	15	Nyaturu	552,344
6	Haya	828,272	16	Asu	529455
7	Makonde	803,974	17	Zigua	436,762
8	Maasai	803,463	18	Kurya	423,511
9	Hehe	740,113	19	Yao	416,802
10	Nyakyusa	723,990	20	Luguru	403,355

Before the late 1980s, Tanzanian researchers could not work on the ECLs, as the emphasis was on nation building and its linguistic corollary - strengthening the national language. Political changes, especially with the multiparty system, have led to the loosening of such restrictions, and local scholars are now moving to fill the gap. This study of the Chasu community also enjoys the fruits of the abovementioned cultural policy that now acknowledges the existence of ECLs and makes provision for research on them.

Nevertheless, like many other ECLs in Tanzania, Chasu has been threatened by Swahili and English due to the multilingual Tanzanian language pattern. Mreta (1998) indicates that recently the functional domains between Swahili and Northern Chasu have become less clear cut. This appears to be the situation, with many ECLs in Tanzania losing ground to Swahili.

There might be some variation within the language itself, but its contact with other languages which are perceived to be comparatively more powerful can result in extensive code-switching and in lexical borrowing. The preliminary fieldwork and further actual fieldwork conducted in respect of this study has given an indication of the existence of these phenomena, hypothesized to be caused by the language contact. Together with phonological variation, this study examines some patterns of lexical borrowing and code-switching as a result of language contact in the Chasu speech community.

1.5 Chapter conclusion

This chapter has introduced the Chasu community in particular and the language situation in Tanzania in general. As already stated, this study focuses on the sociolinguistic variation in relation to social variables in a multilingual rural African setting. This is intended to redress the emphasis in sociolinguistics on western urban monolingual societies or on African multilingual, urban centres. The following chapter presents a conceptual framework outlining the methods and analysis of the study to follow. It then provides an overview of studies based on western urban societies and sociolinguistic studies in Tanzania, and explains how they validate the need for this study in a rural setting.

CHAPTER TWO
CONCEPTUAL FRAMEWORK AND LITERATURE REVIEW

2.0 Introduction

This chapter focuses on three main concerns. First, it briefly presents the variationist paradigm as the sociolinguistic approach mainly used in this study, and, secondly, it reviews some studies of sociolinguistic variation in urban western societies. It also gives a review of the sociolinguistic studies conducted in Tanzania, particularly in relation to language contact.

2.1 The Variationist Paradigm

Sociolinguists have long demonstrated that individuals exhibit variation in their speech in respect of particular features and speaking noticeably differently according to situational context. In every community, regularity in inter-personal and intra-personal linguistic variability has been uncovered. Such differences have occurred at all levels: lexical items, grammatical forms, and especially accent. Variationist sociolinguistics as a quantitative approach principally draws upon the foundational studies conducted by William Labov in New York City and Philadelphia in the 1960s and 1970s. It was developed as a theory that could help to advance the methodology and analysis of language variation and change. It essentially combines techniques from linguistics, sociology, anthropology and statistics to investigate language use and structure. Variationist sociolinguistics seeks to study the balance between linguistic and social structure, grammatical and social meaning. These basic attributes of language refer to both social and structural linguistic factors (Tagliamonte 2006:5).

This approach demonstrates the systematic nature of much of the linguistic variation that was previously thought to occur as a random event or free variation. The

latter is equivalent to saying that there are events whose outcome cannot be predicted with certainty. Variationist sociolinguistics has also been termed "correlational sociolinguistics", as analysts attempted to correlate linguistic variation with language-external factors such as ethnicity, gender, class and age (Trudgill 2002:373). Correlational studies eventually helped to demonstrate the myth of "random variation".

For some years historical linguists and dialectologists explored regularity in linguistic variation by relating it to "natural dimensions", i.e. time and space (Milroy and Milroy 1997:49). The main advances have been in the "human dimensions" of variation, i.e. social variation. The variationist approach employs quantification as its essential methodological tool, which involves counting and comparing the occurrence of variants from different speakers or groups. It is important to understand that social variation was the first aspect of language to be studied quantitatively.

Expressing the fundamental premise of the variationist approach, Weinreich et al. (1968:99-100) as cited by Bayley (2002:117) argue that "an understanding of the language requires an understanding of variables as well as categorical processes and that the variation that we witness at all levels of language is not random". Rather, linguistic variation is characterized by "orderly or structured heterogeneity" or "normal heterogeneity". This implies that language varies as speakers have different ways of saying more or less the same thing within the same language. Heterogeneity is furthermore not random, meaning that the choices that groups of speakers make from within a linguistic variable are systematically patterned by structural linguistic and social factors. The variable patterns both reflect and partially constitute the social organization of the communities to which the language users belong. From these variable patterns, speakers transmit 'cognitive' information, while at the same time making statements about their social identity; what social group they belong to; the kind of conversations they consider themselves to be engaged in and how they relate to their hearers (Tagliamonte 2006:7). These patterns can be studied and discovered only through systematic investigation of the speech community. Labov (1994) also adds that synchronic variation is frequently a reflection of ongoing diachronic change.

With respect to characterising such variation, Young and Bayley (1996:253) describe two main principles which are the "principle of quantitative modelling" and the "principle of multiple causes". The "principle of quantitative modelling" expresses quantitatively the association between the contextual features and the occurrence of the linguistic variable. The linguist as analyst systematically examines the forms that a linguistic variable takes and notes down the features of the context which co-occur with these structures. The context is supposed to be the surrounding linguistic environment in which there is co-existence of social phenomena with a given variable form. Making a statement regarding the possibility of the co-occurrence of variable forms and any contextual features requires a set of large enough data. By the "principle of multiple causes" we mean that multiple contextual conditioning factors can exhibit variability observed in natural language data (Bayley 2002).

The variationist approach is empirical in its methods in the sense that it normally draws upon large bodies of systematically collected naturalistic speech from real speakers. Variations insist upon analyses that reveal full accountability to the data collected and that develop socially sensitive accounts of language change (Labov 1994, Milroy 1997:48-49, Chambers 2003). It is a data-led approach in the sense that patterns observed in the data must be accounted for whether they support some prior theoretical position or not. This is unlike a theory-based approach in which the function of the corpus of data is likely to be lessened to one only of verification rather than being a primary information resource (Docherty et al 1997:279). In addition, to come to an understanding of variation and change in the structural parts of the language, variationists sometimes use insights from the behaviour of the speakers and their activities in naturalistic settings. These can lead to conclusions about varying structures of language and the speakers' knowledge and use of these variable structures.

There are several steps which are to be carried out when doing variationist research. The first task is to define the linguistic variance (Labov 1975:7). This begins with noticing salient linguistic variables; these are linguistic entities which are known in

advance to have different realisations (Hudson 1980:139) based on the social, stylistic and or linguistic parameters (Feagin 2002:23). The linguistic variables are the items which occur frequently in the course of undirected conversations and they should also be structured; implying that the more the item is integrated into a large system of functioning units, the greater will be the intrinsic linguistic interest of our study. This step searches for the largest environment in which this variation occurs in order to apply the *Principle of Accountability.* As formulated by Labov (2005:3), the *Principle of Accountability* requires that reports of occurrence of variants must be accompanied by reports of non-occurrences. It is thus necessary to identify sites where a particular variant can occur and to note the instances of that variant as well as those of other variants of the variable in that site even if the patterns revealed do not immediately support a prior specified theoretical position.

Language analysts should pick out from a text not only those variants of a variable which tend to confirm their argument, but also those which do not, i.e. ones that occur in the same environment as overlapping rather than discrete variants (Lesley Milroy 1987:113). According to Labov (2005:3) "the definition of variables is attained by locating and setting aside *neutralisation*" this is an "environment in which it is not possible to distinguish all variants of the variables". Variationists also pay heed to *exclusions,* where "individual items behave differently from other members of their class". All of this enables linguists to establish the nature of the dependent variables. Examples of the most studied variables include variable (ing) which can result in alveolar [n] or velar nasal [ŋ] in *'ing'* word endings such as *running, singing, walking.* It can be variable (r) which may give rhotic or nonrhotic invariants postvocalically as in words such as *beard, guard, cart* and *yard* (1974a:47, 158). Another variable (th) which is realized in Detroit Black English Vernacular contains four variants which are standard [θ], labio-dental [f], stop [t] or zero [Ø] in words such in words like *tooth, nothing, with* (Chambers 2003:128).

What follows is to locate and define the independent variables. Then are usually social and stylistic constraints on the variables. The selection of these constraints involves

searching for every likely category which can have an effect on the dependent variables in particular ways or include every category that has been found to be significant in previous research. In quantitative analysis of sociolinguistics, these constraints are normally found through the binary division of the population (of speakers or utterances) into salient groups such as men vs. women, children vs. adults, and middle class vs. working class. For instance we may initially make a hypothesis that men and women in a particular society contrast in their use of a particular pattern of linguistic variables and that younger and older members of the same community differ with respect to another pattern. Normally the definition of both linguistic variables (dependent variables) and independent variables can be through the intuitions of the linguist as a native speaker, via prior studies, or perhaps conducting a pilot study (Feagin 2002:21-23). A pilot study helps to substantiate the existence of variables and also narrow down the focus of the project. Where the distribution of variables is highly stratified, a preliminary exploration can suggest an asymmetric distribution across specific strata of the speech population. Where the variable is less obviously stratified, the variationist analysis is more open-ended and exploratory.

The next step is to collect the texts of spoken data (Hudson 1980); this task necessitates looking for representative speakers who are willing to be interviewed and recorded. This typically involves gaining the confidence of a certain group of individuals and obtaining their permission to record them. Demographic information can be obtained through questionnaires which can reveal education level, occupation, type of residence and neighbourhood, demographic features such as age, sex and ethnicity etc. Texts in the form of spoken language can be attained through the interviews, extended conversations, reading word lists, participant observation, rapid and anonymous observation, telephone surveys etc. (Feagin 2002:26-35).

It is essential to find an appropriate recording instrument which will give a sufficiently clear recording, enabling the easy identification of phonetic variants. These days recording usually involves the use of digital recorders like MP3 recorders. To mention a particular type of recording equipment is not useful, because the technology

changes rapidly. The main focus should be to select an instrument which improves the quality of the sound, minimizes the speaker's attention to the recording mechanism, and is easy to use and carry. However, it is seldom possible to record the speaker's most natural speech sound: what the sociolinguist does is to come as close as possible to "the vernacular" as used by community members amongst themselves.

When the recording process is completed data analysis begins. The first stage is to systematise the corpus of interviews, giving labels to each interview and pseudonyms to interviewees. The analyst finds it useful to make notes about the background to each interview, and then the process of listening through the interviews and noting passages for transcription follows. The ideal of transcribing all one's data verbatim is always possible except in cases involving large data bases, for those who are blessed with a research team and funds for transcribing. However, even the less well-endowed linguist has to listen to all the data to get a feel for the community's speech norms. Then the variants of the selected variables can be focused on. This task involves careful attention in recognising the phonetic variants. However, sometimes even highly trained phoneticians can produce different analyses of the same text. The process of identifying the variants may easily be achieved only if there is already a clearly identified hypothesis and aspects of the possible relevant environment.

Data analysis involves counting the identified occurrences of each variant in each text and then comparing the figures of different texts. At the same time cumulative tabs for the identifiable subgroups of speakers can be done. Then follows the task of discovering which differences between the texts are significant; i.e. which would form a reasonable basis for generalising to other texts of the same types. The procedure of calculating these figures depends also on the number of contextual factors that promote or inhibit the use of particular variant. When conducting a study that relates variation to a single contextual factor, simple statistical procedure such as comparing two means with the aid of a *t-test* can be used (Beebe 1977). Sometimes the figures can simply be reduced into percentages for much easier comparison (Hudson 1980). This model can nevertheless be inadequate when multiple influences are likely to be involved in

producing the relevant data. Hence "Analysis of Variance" can be used to relate variation to a single independent variable with multiple levels. In variationist linguistics, however, VARBRUL has been developed as the main computer-assisted tool for performing multivariate analysis. The programme was developed by Sankoff and Rand; and a Windows version was developed by Sali Tagliamonte. The VARBRUL programme has been "proven to be a powerful analytic device for identifying significant linguistic, social, and interactional factors that differentiate or condition probabilities associated with linguistic variables" (Berdan 1996:209). Where interaction among or between non-linguistic and linguistic factors is the main focus of the variation study, other models like the logistic regression model or chi-square are preferable, as VARBRUL cannot serve effectively. In this study, particularly with phonological variation, I have used VARBRUL over other statistical techniques and where VARBRUL failed, a new development, Rbrul which involves the integration of VARBRUL within the statistical computer programme, was applied.

The next step is to interpret the results. This involves fitting the findings into the general theoretical framework dealing with the structure of language and its relation to society and individuals. The linguistic variables are at this stage correlated with the co-occurring contextual features. Correct methodology at all previous stages and setting the adequate general theoretical framework determines the success level at this stage.

In conclusion, research in the quantitative paradigm demonstrates the systematic nature of much of the linguistic variation that was previously thought to be random. For several decades, work within the quantitative paradigm has contributed extensively to the progress of studying language in society, particularly in Western communities, as will be shown in the section below. It has attempted to provide proof for the systematic nature of the variability between linguistic and social variables. It is clearly shown in research that, linguistic variable forms are constrained by multiple internal and external factors. And research has shown that at least with respect to major linguistic constraints, given sufficient data, individual patterns do in fact match group patterns.

2.2 Literature Review

This section reviews in detail what others have contributed to the study of sociolinguistic variation, contact and change. It starts with studies done in Western urban communities, followed by sociolinguistic studies of African languages which were conducted in Tanzania, particularly in relation to language contact.

2.2.1 Western urban communities

The analysis of sociolinguistic variation in Western urban communities has been basically guided by the quantitative approach in methodology and data analysis. Language variation is revealed to be a socially regular pattern which also facilitates an understanding of the mechanisms of linguistic change. It is a socially regular pattern of variation, as it is the result of a systematic examination of the relationship between language variation and a speaker's social variables like social class, gender/ sex and age. Therefore, this review includes those studies in which these extralinguistic variables are interrelated in urban Western communities. In sequence, the sub-sections review studies of language variation in relation to social class, followed by gender and then age.

(a) Social class

The social class to which we belong imposes certain norms of behaviour on us and is reinforced by the people from that group with whom we associate most closely. Most quantitative sociolinguistic studies in Western societies have followed Labov (1966) in initially setting out to examine the relationship between variables of language and social class. Of the principal social dimensions that sociolinguists have been concerned with, social class and gender have probably been the most researched. The patterns of social class differentiation are often assumed to be fundamental, and other sociolinguistic patterns of variation like gender, age, style and network derived from them. Michael (1962) as cited by Ash (2002:402) comments that social class can be considered to be "an individual's life chances stated in terms of his relation to the production and acquisition of goods and services". This has generally been taken to be a

central concept in Labov's (1966) social stratification in his study of the Lower East Side of New York City. Based on Max Weber's (1884-1920) idea of social class, Ash (2002:403) comments that the two main components of social class are firstly "the objective, economic measures of property ownership and the power to control it confers on its possessor" and secondly "the subjective measures of prestige, reputation and status". Nevertheless many sociolinguistic studies which grouped individuals into social classes have followed the generally accepted sociological practice of assigning numerical values to indicators or sub-elements of social status such as education, occupation, income, neighbourhood or type of housing, wealth, family background etc. All these play a role in determining the people with whom we will have daily contact and more permanent relationships. It is from these essential distinctions that we can separate manual from non-manual workers or lower- from to upper-class people (Milroy 1987:100). The typical social class continuum correlates with a linguistic continuum from standard to non-standard or vernacular, with vernacular forms most prevalent among the lower- or working- class people and the standard norms used most frequently by the upper-class non-manual workers (Fasold 1990:224). Ash adds that the dimension of social class can become even more important and productive in sociolinguistic research if it is used more systematically and applied in comparable ways by researchers working in different communities.

Labov was the first to introduce systematic methodology for investigating vowel and consonantal variation in social dialects and in the first large scale sociolinguistic survey of urban community. Labov published works on sociolinguistic surveys such as in Martha's Vineyard, the Lower East Side of New York City, a sociolinguistic survey in New York City within three department stores (Labov 1972:44-49) and the study of language change and variation in the neighbourhood of Philadelphia conducted in 1970s. Since then he has been engaged in a national study, the Atlas of North American English (Labov et al. 2006).

In these works, social stratification was a central variable. In a sociological critique, Williams (1992:80-81) comments that Labov used both subjective and objective conception of class stating, as he quotes Labov (1972:44), that "social stratification is the

product of social differentiation and social evaluation which is the product of the normal workings of the society''.

As mentioned earlier, the subjective conception involves the use of prestige rankings and awareness of status groups who share a similar amount of prestige in a certain speech community. In Labov's study in the New York City Department Stores, stratification was subjectively defined by the prestige of the three large department stores that were studied in Manhattan. The following independent factors were used to set up the relative prestige of the stores: 'the location of the store; the amount of the advertising in the *New York Times*, with its middle-class readership, and in the *Daily News*, a working-class newspaper; the comparative cost of supplies in the stores, the form of price quoted in advertising copy, the relative emphasis on prices; the physical plant of these large departmental stores; and the information on the regard held by the employees for working conditions at the three stores' (Ash 2002:405). Based on these decisive factors, the social stratification of the three stores was firmly established, while in a subtle manner, the study controlled for occupation group: hence the interviewees were predominantly 'sales people' including 'a small sample of floorwalkers, cashiers, and stock boys' (Ash 2002:406). It was speculated that sales people would show stratification comparable to the customers' ranking as top, middle or bottom. Consequently, the social rankings were Saks Fifth Avenue - high; Macy's – middle, Kleins - low.

On the other hand the "objective conception" focuses more upon social differentiation, where the key element is the manner in which differences exist between objective groups, rather than the nature of the status and power relations that exist between them. It involves several dimensions of differentiation, such as occupation, education, income and location of residence. In the study in the Lower East Side of New York City, based on the earlier study in the Mobilization for Youth program (MFY), the scale of socio-economic class was explicitly established relying on objective factors which included the occupational rank, the level of education and the categorization of the family income. These factors were weighted equally in calculating the index score, which ranged from 0-9, resulting into four categories, namely: lower class, working class, and

middle class. The last category was then divided into lower-middle class and upper-middle class. In sociolinguistics, the upper class do not generally feature, as they are always small and inaccessible. Together with education and occupation factors, residence value was added in constructing the socio-economic status index in the work on language change and variation in the neighbourhood study in Philadelphia (Ash 2002:407).

Another important aspect to be reviewed is how to select speakers i.e. sampling. In Labov's (1963) study of Martha's Vineyard, Labov used a judgemental sample; this involves selecting subjects to fill the pre-selected social categories, crisis-crossing geographical area, profession and ethnicity. The pre-selected judgemental sample varies from numerous criteria such as being native-born, parents also being born in the same area, or subjects being a 'friend of friend' (the Milroys 1980) in Belfast. Horvath (1985) also used a stratified judgemental sample in Sydney, Australia, and analysed it using principal components so that the analysis incorporated no assumptions about class membership or sex. The principal components analyses put speakers into clusters according to their linguistic similarities and in that way revealed what the sociolinguistic groupings of Sydney were, based entirely on speech. Unlike the study in Martha's Vineyard, in New York City, Labov selected informants by means of a scientifically designed random sample. By "random sampling", it is meant that though not everybody in the population is interviewed, everybody has an equal chance of selection for interview. This technique enables the informants to form a truly representative sample of a particular area of investigation. If the informants are a representative sample of the linguistic description, then it could therefore be an accurate description of all the varieties of the language spoken in a particular area of study.

Commenting on Labov's methods, (1974a:39) argues that the methods have provided evidence that has proved vital to the study of social-class dialect and accents. Labov has shown also that, viewed against the background of the speech community as a whole, variation in language is not free or random, but determined by extralinguistic factors in a predictable way. Rapid and anonymous observation works best as a supplement or a preliminary to other methods. In any case the methods employed by

Labov partly addressed the problem of heterogeneity of speech communities. This has helped sociolinguistic researchers to correlate linguistic features with social class accurately, and obtain thereby a clear picture of social dialect differentiation (Trudgill 1974a:39-40).

In the study of English in Norwich, England, where Trudgill (1974a) investigated the co-variation of phonological variables and social differentiation, the construction of a social index was based on multiple measures (Ash 2002:410). This included parameters such as occupation, father's occupation, income, education, locality and housing as well as four contextual styles. From these social parameters and stylistic stratification the mean score for each social group was calculated for the variable (ng). The groups were; lower-working class (LWC), middle-working class (MWC), upper-working class (UWC), lower-middle class (LMC), middle-middle class (MMC), word list style (WLS), reading passage style (RPS), formal style (FS) and casual style (CS). This is not substantially different to Labov's procedures.

Though Trudgill declares that occupation is not a critical factor in arriving at this kind of grouping, since its weight is only two-fifths of the index (own occupation and father's occupation), he nevertheless finds that the occupational range of each class is highly systematic. (Ash 2002:410-11) argues that the highest class (MMC) that Trudgill describes for England consists mainly of professional people, including teachers, managers, employers, bank clerks and insurance workers. The second group (LMC) is made up of non-manual workers such as typists, commercial travellers, and office workers. The third (UWC) consists of foremen and skilled workers'; the fourth (MWC) includes manual workers; and the last and lowest is mainly made up of unskilled workers. Through these means, the main three objectives were attained, including:

> (a) to investigate the nature of the correlation between realisations of phonological variables and social variables, (b) to discover which variables are subject to social class differentiation and which to stylistic variation and (c) to find out which variables are most important in signalling the social context of some linguistic interaction, or the social class of a speaker. Trudgill 1997:179

Ash (2002:411) comments that, while other sociolinguistic researchers do not take occupation by itself as a sufficient determiner of social class, Horvath (1985) used it alone effectively in her study of variation and change in Sydney. On the basis of occupation alone she managed to rank three classes. These were (a) middle class, which consists of professionals and skilled workers such as accountants, real estate agents, and pharmacists; (b) the upper-working class which consists of less skilled workers like attendants, arc welders, builders, chefs and salesmen, (c) the lower working class, consisting of unskilled workers such as truck drivers, metal workers and factory workers. Ash (2002:11) has cited Lennig (1978) as using occupation alone to categorise the social class in his study of variation and change in the vowel system in Paris, while Haeri (1997), in studying Cairene Arabic, used occupation and educational background.

Generally, through the introduction of these methods of investigating social dialect by correlating sociolinguistic variables with social factors, sociolinguists have been able to construct a comprehensive representation of social dialect differentiation in the Western world and in other places where these studies have been replicated. It is undeniable that the dimension of social class is not only important, but is also highly productive in sociolinguistic research if it is applied systematically and in comparable ways by researchers working in various speech communities. However, although many sociolinguistic studies investigate social class differences in linguistic variation and change, social class itself is a difficult concept to define, quantify and interpret, particularly where children and women are concerned (Rickford 1986, Ash 2002:20, Milroy and Gordon 2003). Many sociolinguists have conceptualized social class based on a combination of features like stylistic context, housing, education, income and occupation, but the single factor that accounts for by far the greatest portion of the variance is occupation. Ash (2002:419, 20) declares that it is difficult to envisage a composite index of social class that excludes occupation, and some researchers have even used occupation alone in defining social class. In the industrialised world occupation prestige is considered to be fairly stable over time, at least in a short run, except for some changes which may be caused by the ‘prestige of scientific occupations, a decrease for

culturally oriented occupations, and an upward trend for artisans' (Hodge et al. 1964 as cited by Ash 2002:419). The finding that occupational prestige is stable is a result of presumed stability in the prestige connected with the criteria on which the ratings are likely to be based, such as education, income and functional importance. Factors other than occupation which are indicators of social class can also be used in a motivated way, with an awareness of the distinction between objective factors relating to economic power and ownership as opposed subjective factors which include matters of status and prestige.

(b) Variation according to gender/ sex

It has been taken for granted by many followers of Labov that socio-economic class should be treated as the most significant speaker variable. This also implies that other social variables like gender, age, network and social mobility are then interpreted as being attached to or dependent on social stratification (Milroy and Milroy 1993:57-58). Other researchers, however, believe that other social variables, particularly gender, cannot always be accounted for primarily in terms of social class. It is important to note that in various communities, there is social division of labour for men against women in different ways, with demonstrable consequences in the use of their local language. Coates (1986:57-78) has also regraphed a substantial amount of data from several sociolinguistic surveys, which also demonstrate that the sex of a speaker quite frequently accounts for patterns of variation at least as well as, and in some cases better than, social class. Generally, gender-marking in linguistic varieties arises because language is a social phenomenon which is closely related to social attitudes and social economic roles. Men and women are socially unlike, as the society sets down diverse social roles for them to perform and expects contrasting behaviour patterns from them. Language variation and change are a reflection of this social fact (1974a:94). Hence it is unsurprising to find sex-preferential variation, where, for instance, women in a certain community use one variant more frequently than men.

In quantitative sociolinguistic research there is a general finding from Western urbanized societies that despite their social status or age, women tend to consistently use

a higher proportion of standard variants or prestige patterns and style shifting than men (Trudgill 1974a:91, 93; Milroy 1980:112, Milroy 1987:102, Fasold 1990:92, Romaine 1994:78, 100). Women are sometimes thought to initiate linguistic change (Labov 1972a) and to abandon older linguistic features faster than men. In general Milroy and Milroy (1993) have suggested that it is misleading to say that females favour prestige forms, rather they create prestige forms in the sense that those forms they prefer to use become overtly prestigious forms in the society. It is important to note that researchers usually take one of the variants as standard or overtly prestigious, usually on the grounds that this variant is increasingly used in more formal speech style (Cheshire 2002:425).

The general finding that women tend to use a higher proportion of standard variants compared to men of the same social class has been taken in some textbooks as a deep-seated principle of sociolinguistics. Nevertheless the area is a controversial one. Coates (1986) has pointed out that no satisfactory explanation has emerged as to why women should be more oriented than men to a prestige norm.

Generally Labov and Trudgill established similar frameworks for characterising gender differences. Labov tries to resolve two slightly contradictory questions as to why men use more non-standard varieties and why women lead in language change. Labov (1990:214) iterates that "women … are said to be more expressive than men or use expressive symbols more than men or rely more on such symbols to assert their position" and "women are said to rely more on symbolic capital than men because they possess less material power".

However, Trudgill's (1972, 1983b) approach is not very different to Labov's. In his work in Norwich he observed that men use more non-standard forms than women. As to why working class men (and young women in his study) stick to non-standard forms, Trudgill uses Labov's notion of "covert prestige" to explain that men's non-standard variants serve as solidarity markers which highlight certain group values like masculinity. Like Labov, Trudgill (1972:91) states that "the social position of women in the society is less securing than that of men… It may be… that it is more necessary for women to

secure and signal their social status linguistically". Trudgill (1983a:167-8), Deuchar (1988), Fasold (1990:93, 95), Romaine (1994:79) and Holmes (1997:135) suggest that women in the society are denied full opportunities in education and occupational roles etc, as the traditional norms place them in an inferior social position. They need to set a strategy for maintaining face interactions in situations where they are considered to be powerless. So where men can attain social power through their occupational status, earning power and possibly their abilities, women can explicitly secure and signal their social status via language, especially through their use of the standard, overtly prestigious variants. Trudgill goes on to argue that in his study in Norwich, women tend to exaggerate their actual usage of standard forms while on the contrary men tend to underreport their standard usage.

Based on the consistent research results of more than three decades of sociolinguistic research in the speech community concerning the linguistic differentiation between men and women, Labov (1990:205, 210-15 and 1998:7) outlines three principles below:

Principle I. In stable sociolinguistic stratification men use a higher frequency of non-standard forms than women.
Principle Ia. In change from above women favour the incoming prestige forms more than men.
Principle II. In change from below, women are most often the innovators.

Principle I thus suggests that women tend to favour the variants with overt social prestige while men do the opposite. This has been presented in some sociolinguistic literatures as a fundamental precept of sociolinguistics. The overwhelming majority of variable studies have shown this effect not only in urban, industrialized western societies but also in isolated rural areas and large cities in Caribbean (Lopez 1983), as well as in various languages of South America (Albo 1970). A few studies cited by Labov (1998) have also shown that not all sociolinguistic variables mark sex effect. Several Tokyo variables which Hibiya (1988) studied indicated no significant sex differences. The same

applies to Morale's study (1986) of the velarization of /n/ in Puerto Rican Spanish. Very few cases are reported where men appeared to favour prestige forms more frequently than women, of which most are in Muslim societies. Abd-el-Jawad (1981, 1987) reports that, in Amman (the capital city of the Hashemite Kingdom of Jordan) and in Nablus-Palestine, men of all social classes favoured the prestige variants of the variable /q/ more frequently than women. In Tehran women in all social classes happened to use colloquial forms of variables (an) and (æš) more frequently than men (Modaressi 1978). This can be due to lack of accessibility to the prestigious forms, as some women in the Muslim world perform fewer roles in public life. Therefore Principle I is applicable only when women have access to the standard patterns and the opportunity to perform or apply them is also presented (Labov 1998:14). James (1996:119) concludes that there is far too much variation across and within different communities for any simple analysis to be viable. She comments that local economic conditions, the employment and education opportunities available to each sex, social conditions affecting network strengths, the amount of status and respect accorded to women in particular communities and the extent to which they participate in public life are just some of the factors that may account for the choices that women and men make in the speech features they use. Labov's principle I has further received more challenges as Milroy (1991) argues that the notion of standard and non-standard and the related notion of prestige are not uniform in all communities, and are hence not to be taken for granted. The concepts may carry different meanings to different communities; different groups in communities or even in the same community, as well as in lives of different individuals.

Another challenge is related to assigning women to different social classes in the hierarchy. In her discussion of problems associated with the patriarchal concept of social class, Romaine (1999:174) notes that where the family is the basic unit of analysis, the man is regarded as the head of the household, and his occupation determines the family's social class. Recently, women have been classified in terms of their occupation, though several problems still persist, especially with women who don't have occupations outside their homes. This implies that it is difficult to compare men and women as they do not have equal status within and outside the household (Eckert 1989: 255).

In Principle Ia women tend to lead in the acquisition of new prestige patterns and the elimination of stigmatized forms. These patterns are composed of the changes from above, which take place above the level of social consciousness mostly in formal style, essentially adopted from norms external to the speech community, and are often subject to hypercorrection (Labov 1998:14). Based on the empirical evidence from the Philadelphia survey, Labov (1990:240, 1998:21) insists that, as a rule, women are active innovators of change, particularly those of the lower-middle class. He hence concludes that "the interaction of sex and social class leaves us no choice but to focus on women's behaviour and to assess its effects on linguistic change" (Labov 1998:39). However, in other places this involves shift from one language or dialect to another as a result of economic factors. In this case, language is associated with the work situation and educational opportunities, which are predominantly open to males, as in Papua New Guinea (Sankoff, 1980:123). Hence women cannot be expected to favour the prestigious forms since they cannot access them. Therefore in such a case men will dominate in language change. On the other hand, in unprivileged communities, sensitivity to exterior standards of correctness in language is associated with upward mobility. Female students in inner city black communities have been more successful than males in school and in employability partly because of their greater accommodation of the external standard form of American English.

Labov (1998:15-17) comments that Principle II involves changes from below which are the basic forms of linguistic change functioning within the system below the level of the conscious. Women are considered to lead in this kind of change, while men are conservative in respect of the internal variables of the speech community. The study of sound change in progress, done by Trudgill (1974b), in the backing of (el) in *belt, help*, etc. in Norwich, showed women to be ahead of men. Women are innovators in the development of a new norm for (aw), a fronting that was added first to and then substituted for the traditional centralization before voiceless finals (Chambers and Hardwick 1985) as cited by Labov (1998). However there are minor counterexamples where men have been shown to be ahead of women. The centralization of (ay) and (aw)

on Martha's Vineyard (Labov 1963) and the unrounding of (o) in Norwich (Trudgill 1974) were male-dominated changes.

These tendencies lead to the conclusion that it is difficult to connect the mechanism of change to sex difference because either sex can dominate the change. Labov (1998:43) concludes that, from the linguistic point of view, Principle II is more challenging, as it is less regular than Principle I, possibly because there is no way to ascertain in any given case whether men or women lead at the beginning of a linguistic change. However, in the entire discussion on sex differentiation in relation to language variation and change, women have frequently been found to dominate in change ahead of men and the number is relatively small in isolated cases where men favour more standard language and they are the innovators of linguistic changes. As mentioned earlier, in other societies where acquisition and usage of language are associated with economic power and education is easily accessed by men, women may use less prestigious forms as they cannot access the standard style.

(c) The interaction between sex/gender and social class

Eckert (1997:212-226) explains that in the study of language variation and change there has been a tendency to detach sex/gender from social class. At the same time, socio-economic class has gained more attention in the sociolinguistic literature compared to gender, which has suffered from being given less scientific attention. More attention has been given to sex as a biological and demographic category rather than the roles, norms and expectations of society which constitute gender, a social practice and a social construction of sex. Gender roles in the community involve more than merely competing with a member of either sex category; rather, men may perceive their social status in relation to their fellow men, and women will do the same to other women. Eckert explains how boys and girls in Detroit follow different routes in achieving power and status within a membership group. The study is about two social groups which dominate adolescent social life in American public high school, the 'Jocks' from the middle class and the 'Burnouts', a group from working-class culture. While boys in these groups maintain their power by showing what they can do, girls maintain their symbolic

difference by showing what they are primarily with respect to other girls. This is highly reflected in their linguistic difference. Eckert explains that adolescents who constantly use phonological variables are those who show a great amount of symbolic work to affirm their membership in their group. Women are constrained to show more of their personality, social qualities and membership, which is reflected in their linguistic variation. From this argument Eckert emphasises the importance of examining linguistic variation within one sex in a social category before comparing it with the other sex.

In the community where gender roles are separate, men and women will compete for power, access to property and services to establish their status in the role domain of the other sex. Where men establish their status in terms of accomplishments, possessions and institutional status, women mostly use symbolic resources to establish membership and status to gain upward mobility, especially where hierarchy happens to be salient. These behaviours are eschewed in the socio-economic hierarchy and hence necessitate an integration of social category and sex as interrelating variables when examining the language variation.

As he examines the interaction between social class and sex, Labov (1972) argues that the tendency to avoid stigmatized varieties and prefer prestige norms is greatest for women of the lower-middle class, and is often minimal for the lower class and upper-middle class. The high tendency towards hypercorrective behaviour among the lower-middle class women has for some time been considered a significant ingredient in the mechanism of linguistic change. In the study of (ng) in Norwich, Trudgill (1974b) claims that social groups preserve the expected relationship between male and female in both casual and careful speech, while lower-middle class women dramatically cross over the lower-middle class men from casual to careful speech. These incidences of interaction have strong implications for our understanding of the role of women in sociolinguistic variation. Most often women in this group have considerably more political and economic power than the working-class or lower-class women, and are capable of making more money and grasping more opportunity than their male partners in the upper-working class.

On the question of the interaction in linguistic change from below, most substantial findings have revealed that linguistic change from below is largely follows with a curvilinear pattern (Labov (1998:27, 39). The study in Philadelphia verifies that women, as agents of most linguistic changes, are the active agents of sexual differentiation, particularly the lower-middle class women who have revealed a very different slope of style shifting from all other groups. Women in the lower-middle class have instinctively established the difference between themselves and men as they quickly adopt new prestige forms and brusquely react against stigmatized features.

(d) Age in relation to language variation and change

As mentioned above, language variation can constantly mark class difference or stable sex difference in communities, but it can occasionally mark instability and change. The prime independent social variable to be correlated with instability and change is age (Chambers 2002:355). The effect of age in sociolinguistic patterns is very evident when comparing the speech of adults with that of children. It is known that the differences in anatomy and physiology or biological age are largely responsible for these differences, but socially–oriented variation also occurs in the course of life. In discussing such variations, especially in complex modern societies, at least three stages are taken to give significance consequences (Horvath 1985, Eckert 1997). This is from the formation of local peer groups in childhood through the network of adolescence, to the style of life in adulthood.

The early childhood period is characterised by relatively immature verbal patterns due to ongoing language learning and incomplete growth of the child's vocal anatomy and motor control. However, local forms of pronunciation, which include relatively complex patterns of allophonic distribution, surface from this stage of the acquisition process (Roberts and Labov 1995, Roberts 1997a &b, 2002). Children, from their earliest stages of speech, develop sociolinguistic competence, and they engage themselves in complex register variation and become acutely aware of the relation between social roles and language variability. They learn the social functions of variables before they develop linguistic constraints and use the variables on occasion perhaps as conscious markers in

particular lexical items. In Philadelphia, Roberts' work with three-year olds has shown that children's language at the early stage is inherently variable, much the same as the speech of the older people that serve as their models (Roberts and Labov 1992, Roberts 1993). Three-year-old children have shown variation in their use of both stable variables (such as *–ing* and t/d deletion) and patterns of local variation representing change in progress (such as the raising of short /a/ in Philadelphia). In the speech of children aged 4, 6 and 7, Labov (1989) found adult-like linguistic constraints in t/d deletion and *–ing***.** In most societies where there is a typical pattern characteristic of adult women speech, it is presumed that children have a greater chance to acquire the dialect based on that linguistic input, since women are often primary caregivers. However adults are not the children's only primary linguistic models. Their interaction with siblings, neighbours, and friends exposes young children to changes in progress as manifested in the speech of their older peers, and affords them the kind of participation needed to understand the social meanings of those changes (Hockett 1950). Sometimes they may find themselves using the speech of their age mate regardless of their social class membership, until they become responsive to the language of their social class. Then they start using it.

The adolescent period also reveals the linguistic influence of the peer group and sometimes its role can overtake the influence of the home (Eckert 2000). By the age of 12 the pressure to conform to peers' norms is great enough to eradicate most of the initially-acquired linguistic difference. This leads to a strikingly homogeneous local accent. It is the time when linguistic change from below is advanced. In other societies, adolescents lead the entire age spectrum in sound change and in general use of vernacular variables. Adolescents often engage in constructing identities in opposition to or at least independent of their elders, though sometimes they may conform to the social class norms of their parents. In the study of African-American English in Detroit, Wolfram (1969) as cited by Eckert (1997:163), found perfect correlation between children and adults, but less perfect correlations for adolescents.

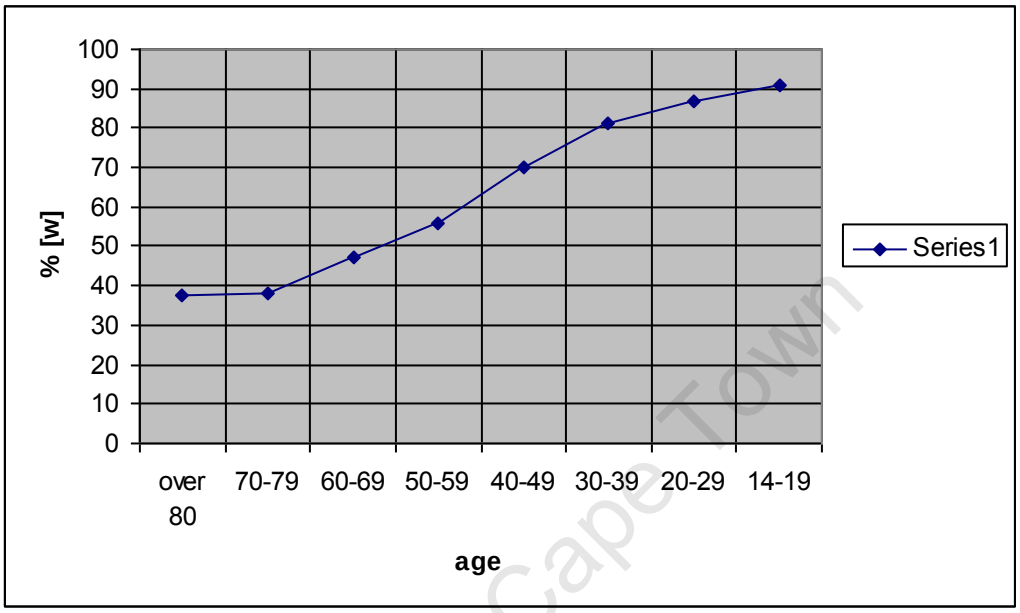
Adulthood is assumed to be stable, with the phonological structure of the language having become fixed. Adulthood is inclined to be a conservative, period as

adults tend to use more stable, earlier-learnt variants than younger age groups, due to pressure for the use of standard language in the workplace. While linguistic conservatism has been the main characteristic attributed to adults, sometimes due to personal social ambitions or the circumstances of the speaker, ongoing changes and deployment of sociolinguistic variants can be marked during adulthood. This can include learning a new dialect or language after geographical relocation (Chambers 1992, Coupland 1980, and Mees & Collins 1999) or changing social status. It is also speculated that adults relax their conservatism after retirement and their linguistic behaviour changes due to a lack of power and pressure associated with the work place. However Eckert (1997:165) claims that this concern has not been explicitly studied.

In connection to age difference, linguistic change in progress reveals its prototypical marks in a pattern when a particular minor variant in the speech of the oldest generation occurs with greater frequency in the middle generation and with still greater frequency in the youngest generation (Chambers 2002:355). Generally a true incoming variant will be marked as a linguistic change by its increasing frequency down through the age scale of younger people.

The study on the use of [w] instead of [hw] in words like *which* and *whine* in central Canada is a good example of a steady and clear progressive change. Graph 2.1 below indicates that there were times when the use of [hw] was stable between age 79 and 80 with some minor changes in the use of [w]. At certain times both the use of [w] and [hw] was almost equal at age 60. Then changes occurred in every decade to favour the use of [w] down to the younger generation. From age 20 down to the teenagers, there has been almost total elimination of [hw] variant and the use of [w] has started to stabilize. This line graph gives an S-curve which is characteristic of linguistic change as it represents different stages of the initial period, the rapid rise and the tailing off stage.

Graph 2.1: Percentage of speakers with [w] not [hw] in words like *which* and *whine* in Central Canada



Source: Chambers (2002:360)

It is evident that age difference marks the characteristics of language change throughout generations. However other factors may contribute to language change. Milroy and Milroy (1985) argue that loose networks and weak ties act as a conduit for linguistic change, since they increase the chances of the exposure to external linguistic patterns. This exposure will make the community susceptible to changes originating from outside localized networks. It is through this exposure that linguistic innovators are likely to contract many weak ties; as a result they succeed in weakening the stable localized community norms.

From these reviews one can conclude that in every speech community linguistic variables tend to be distributed throughout the population, yet are so graded that under more constrained circumstances, particular social groups or classes use particular variants infrequently. It is apparent that variability is within everyone's experience of using and reacting to their languages. Yet in some societies little attention has been paid to variation, or such variability that is acknowledged is considered unimportant, accidental,

or inessential. In this section, English examples are used to illustrate general principles and the problems in the study of phonological variation. However modern linguistics is so dominated by work on English that much of what we know about variation per se is derived from English data and especially from urban North American and British varieties. A great deal remains to be learnt about varieties of English elsewhere and about variation in other languages like those in rural Africa.

2.2.2 Review of Sociolinguistic Studies in Tanzania

Sociolinguistic research done in African and Asian communities has paid more attention to language planning and policy than language contact and variation. In Tanzania, the focus is almost entirely on the national language Swahili, which is also one of the official languages. Special emphasis is given to education planning as the medium of instruction in primary schools and as a subject in secondary schools and higher learning institutions. The influence of Swahili on the ECLs of Tanzania has been the topic of several sociolinguistic studies. Some studies have been conducted on code-switching in school communities. Several studies are reviewed in this section, which will help to place the current study in a national context.

One of these sociolinguistic studies is that of Mochiwa (1979), which investigates the impact of Kiswahili on ECLs. The focus was Kizigua, a language spoken in Handeni district, Tanga in North Eastern Tanzania. The study investigated the language knowledge, patterns of language use and the attitudes of people towards Swahili and Kizigua. A sample of 62 primary school pupils responded to Mochiwa's questionnaires. In his findings he argues that although the Wazigua are non-migrants to Handeni District, his study of bilingualism patterns across age groups shows that they are shifting from the lesser-developed language (Kizigua) to Swahili. The study thus indicates that bilingualism in Handeni is unstable. Interference of a cultural and semantic nature, with lexical borrowings mostly from Swahili to Kizigua, seems to be the most serious effect of language contact on the part of Kizigua of Handeni.

Rubanza (1979) also undertook research which focused on patterns of language use and the attitudes of people towards Swahili and ECLs. In this study, Rubanza aims to assess the relationship between Swahili and other African languages, particularly Kihaya. Kihaya is spoken in the north-western part of Tanzania in the Kagera region. Unlike other studies, which were conducted in rural speech communities, his interest was the language use of rural migrants in Dar es Salaam, focusing on Kihaya speakers currently living and working there. He was interested to find out which areas of communication were still being served by Kihaya and which were in a strong position to resist or to absorb the growing pressure of Swahili. In this study he also investigated how factors such as knowledge of language, education and occupation, topic and environment operate as regulators of language choice. The study reveals tendencies towards language shift to Swahili, even in the domains like home and family, in which Kihaya is the language choice expected predominantly among the youth. Although this tendency was not previously tolerated, it was increasingly becoming acceptable. There is also a predominance of code-switching and interference in terms of lexical borrowing from Swahili to Kihaya but not vice versa. This indicates the possibility, if not the likelihood, of language shift from Kihaya to Swahili among the Wahaya, particularly those living in the city centres.

Another avenue of research was conducted by Mekacha (1993) on the sociolinguistic impact of Swahili on the ECLs in Tanzania among the Ekinata group of Motukeri Village in the Mara region. In this study Mekacha aims to find out and explain whether factors such as language knowledge, language learning, language use, code-switching and communicative proficiency indicate shift from the Ekinata to Swahili. The findings reveal that there is indeed a gradual language shift taking place among the speakers of Ekinata to Swahili. Ekinata predominates among the old people and dominates informal domains (such as family, especially for older family members, neighbourhood and the market place), while Swahili dominates in the domains of religion, government and school and for all domains among the younger generation. Mekacha also emphasized that the contact of Swahili and ECLs has always been vertical rather than horizontal, as Swahili has always been associated with power, modernity and

dynamic life. The study also indicates that the more the community is embedded into the matrix of the wider community, the greater the use of Swahili is.

Unlike previous studies, Mekacha seeks to present the language profile of Tanzania in general, as it is anticipated to be diglossic. However, in the village where the study was undertaken no pattern of diglossia was found, due to language shift. His study shows the difficulties of finding a general language profile for Tanzania.

Msanjila (1999) also contributed to the genre of studies pertaining to Swahili and ECLs in Tanzanian rural areas. His study was conducted in eight rural villages categorised into two main groups, namely 'Near Town Villages' and 'Far Remote Villages' for comparative purposes. The study had three goals: (a) to identify the uses of Swahili and ECLs in rural areas; (b) to investigate in which domains Swahili is mostly used and which domains are reserved for ECLs and (c) to investigate the implications of the co-existence of Swahili and ECLs on the future of ECLs. His study revealed that there are considerable differences in the use of Swahili and ECLs between the "Near Towns" and 'Far Remote villages'. The majority, including young and middle-aged people, female and male literate people, workers in employment and petty traders in 'Near Town Villages' are mainly using Swahili, even in informal domains which were traditionally reserved for ECLs. On the other hand, most people especially the old and the non-literates, in remote villages are using ECLs in most domains. Moreover the continued co-existence of Swahili and ECLs in the rural areas has two implications: (a) disappearance of ECLs in the 'Near Town' villages due to a number of sociolinguistic factors, and (b) the continued survival of ethnic languages in the remote villages. This implies that while ECLs are in danger of disappearing in future (language death) as a result of the current process of language shift in 'Near Town' villages, ECLs in the 'Far Remote' villages are currently least threatened by language shift. However, factors such as improved transportation systems, a positive attitude towards using Swahili, intermarriage, an increase in levels of education, and an improved economic system (changing from peasantry to wider economy market) make the future of ECLs uncertain.

Msanjila categorised the informants into social groups according to variables like age, sex, marriage status, occupation and levels of education. However, these variables were included as demographic information, and not analysed and discussed in detail. Unlike other studies, this one involved more than one ECL in order to arrive at a possible generalization about the national language situation.

Another research project by Kanyuma (2005) investigated social motivations for code-switching in a primary school in Rabuor ward in Tarime district in northern Tanzania. The study examines the use of Dholuo in relation to Swahili and English within a school community in a rural locale. It involved conversations between teachers and their students in using Swahili, one of the languages used as a medium of instruction, and Dholuo, the ECL which is assumed to be used by almost all students when they are out of the school community. The study revealed that conversations between students and their teachers involve code-switching between Dholuo and Swahili motivated by these factors: (a) lack of appropriate lexicon, (b) nature of the topic of conversation, (c) discourse marking, (d) checking for understanding, (e) marking group identity and solidarity, (f) the degree of formality of the context and (g) the need to give concise and succinct explanation.

This is among few sociolinguistic studies that involve variables like age and status (student-teacher relationship) to examine the motivations underlying language contact and variation.

2.3 Chapter conclusion

Sociolinguistic analyses of variation in urban and western communities have decisively revealed language variation in phonetics and phonology from different social groups to establish and/or symbolize socio-economic stratification (e.g. Labov 1972, Milroy 1980 and Chambers 2003). Generally speaking, the findings of urban sociolinguistics do not seem to resonate for non-western communities, especially the non-urban and multilingual communities (Mesthrie 1996). Gumperz and Blom (1972) and Myers-Scotton (1993) have proposed that multilingual communities switch codes in places where urban monolinguals tend to switch styles.

It also appears that sociolinguistic studies in Tanzania focus mainly on language shift and maintenance. The studies which were conducted on the ECLs in relation to Swahili language use indicate gradual language shift from ethnic languages to Swahili. However, a significant number of people in some rural communities may still be using ECLs, as this thesis will show. There is little emphasis in the literature on the sociolinguistic analysis of a language in a rural speech community on its own terms. Sociolinguistic studies in a typical rural-based setting have not been thoroughly undertaken, especially with a focus on the correlation between sociolinguistic variables and social factors. Thus the purpose of the study at hand is twofold. Primarily, it investigates what impact social stratification has on language variation in a rural African community, viz. speakers of Southern Chasu dialect. This is the first systematic research in rural Tanzania which empirically investigates how well-known social variables operate in a rural-based setting in relation to language variation. Taking into consideration that this society is shifting towards multilingualism, the impact of language contact is examined in relation to the given social variables. The focus in chapter 4 is on the structural aspects of lexical borrowing and the syntactic and morphological aspects of code-switched constructions. In chapter 5, the study also examines whether language variation related to social stratification in western urban communities is similar or close to language variation in a rural African community. This is done by testing the patterning

of phonological variables like (s) and (z) which have variable realisations, in relationship to social variables like social class, education level, age, gender and style.

While in Western societies the Labovian model of the sociolinguistic interview is mainly used to gather data for phonological variation, the following chapter shows that this technique serves both objective investigation of phonological variation, and variation arising from multilingualism and language contact.

CHAPTER THREE
RESEARCH METHODOLOGY

3.0 Introduction

This chapter presents the research methods and procedures utilized in this study, the design, and target population, instrumentation and methods of data collection, statistical analysis and presentation.

3.1 Research design

The purpose of this study is to investigate variation in language use, with respect to social stratification. The focus is largely on phonological variation of (z) and (s) but the other facets of variation in a multilingual context like lexis and code-switching were also considered. The research for this study is conducted in Same District, particularly Myamba ward which is comprised of three villages namely Goha, Kambeni and Mangá. The research is conducted in two phases, the preliminary research and the full scale study.

The pilot research work was done during two weeks in December 2006, with the primary intention of exploring how best to undertake the full-scale research and whether the hypothesised variation was indeed discernible. In order to capture an initial sample of data, tape recording was extensively used for the elicitation of oral data.

During the pilot research work, a total of 7 informants were selected, based on socio-economic criteria i.e. occupation and social demographic variables such as age, gender and ethnicity. This pilot sample comprised of 4 males and 3 females with an age range of 50-95. In terms of occupation there were 2 professionals, 1 pastoralist and 4 peasants. Three informants had no schooling at all, 2 had primary school education up to grade 4 while 2 had attended secondary schools and held professional certificates.

The initial module requested the historical background of the interviewee; questions covered the name of the particular respondent, clan and the origin of the clan, the daily socio-economic roles, the ruling system of the clan, worshipping and cultural initiations. The aim of this question was to instigate conversation that may uncover the background data essential for analysing the language variation.

The preliminary study helped to fulfil a range of important functions. In general it assisted in doing the assessment of a full-scale study. Basically, the exploratory phase helped to identify the potential logistical problems or problems in eliciting and analysing the data. It also helped to form an opinion about the socioeconomic groups that exist in the study area, a view which was useful in facilitating the categorization of socioeconomic groups during the actual field work. The conversations held suggested salient aspects of the language variation according to social variables like gender, age, social status and social economic classes. It is from the exploratory study that the major phonological variables for this study were identified. The pilot study also revealed high tendencies of lexical borrowing and code-switching (between Chasu and Swahili) by certain groups of individuals in the community, while other groups seemed to use only Chasu.

Based on the observations of the pilot study, the actual research work investigated whether there are other ways in which a rural community signals stratification besides code-switching. This was done by interviews and subsequent testing of the occurrence of key phonological variables like (s) and (z) in relationship to social variables like education levels, style, social classes, age and gender. Since the society is shifting to multilingualism, lexical borrowing arising from language contact was also examined simultaneously with the structural aspects of code-switching in relation to the given social variables.

3.2 Target population

All the informants for the study at hand are native speakers of the Southern Chasu dialect of Same District. The informants are individuals who were born, raised and educated at primary schools in the same area.

The full-scale research work obtained 57 informants for data collection. This involved 4 social economic classes of (a) peasants, (b) pastoralists, (c) middle-class professionals and (d) businessmen/women. Category (c) of professionals was further divided into (i) political and religious leaders and (ii) teachers and nurses. The informants ranged from the age of 20-85 years. Table 3.1 represents the distribution of relevant characteristics of the informants. This division of the informants is based on the researcher's judgement after observing the society and considering the definition of social class given in section 1.3 of chapter 1.

Table 3. 1 The distribution of the informants

Occupation	Gender		Age range	Total
Businessmen/women	F	07	21-50	14
	M	07	31-60	
Middle-class Professionals	F	07	21-60	15
	M	08	21-79	
Peasants	F	12	31-70	21
	M	09	31-85	
Pastoralists	F	–	–	06
	M	06	31-80	
Total				**57**

In this study, judgmental sampling was employed based on the judgement of relative number of people from different categories in the society. The age groups of the individuals, traditions and customs of the society have a close relationship with the

distribution of the socio-economic activities in the Chasu community. It is explained in chapter 1 that the Chasu community is traditionally patriarchal; hence the ownership of land, houses and animals is titled to men. This is why it is only males who are pastoralists in this data. Individuals below 20 years are still in schooling, so they were not included in this data. Customarily, a young man has to marry in order to claim a portion of land from his parents' possessions. Otherwise he will cultivate the land with his parents and the farm products will be the possession of the parents, as the son is still a junior member of the family. In this study peasants are above the age of 30 years and they are all married. Peasants below 40 years are schooled up to grade seven while those above 50 are schooled up to grade 4 or 8, as they lived during the time of colonialism when the primary school system was extended to grade 8. Four informants did not attend formal school at all, even though one of them is able to read and write. Unlike the peasants and the pastoralists, businessmen and women as well as professionals range from 20 years and above. Except for 3 females who acquired secondary school education and 1 male with a Diploma in business, all the businessmen and women are primary school leavers.

Due to economic improvements considered to be supported by the accessible transport and communication system, youngsters engage in business right after completing their primary school. The exceptions are the few who choose to go for secondary school education based on affordability. This is why it was easier to enlist younger participants from 20 years on in these groups. In the society business is considered to be practised by young people, as they can manage the hardship of travelling from one place to another. In this data, businessmen and women are below 50 years except one subject who was in middle age: 50-60 (see section 3.4.2 a for age group categorization). Apart from 2 informants who had obtained bachelor degrees, other professionals had attained diplomas and certificates after their secondary school education. In this study, professionals cut across all the age groups while peasants and pastoralists do not. This will be kept in mind at the stage of data analysis and interpretation.

3.3 Techniques of data collection

The technique of gathering data which is adopted by researchers is contingent upon what the research is geared at, the context where the information is gathered, sufficient time and financial capability. It is meaningful for phonology-oriented research to use carefully scripted texts and wordlists, because the same words carrying phonological variables or contrasts will be articulated by every subject (Feagin 2002:26, 33). It is constructive to make use of sociolinguistic interviews for the frequently occurring variables, especially phonological and morphological and certain syntactic forms. But it is often necessary to employ participant observation to collect supplementary materials which may not be well captured in recorded interviews (Feagin 1979). When studying phonological variables, sociolinguistic interviews are more fruitful if they are supplemented by more formal styles like reading passages, wordlists, minimal pairs (Labov 1966 & 1974b) etc. Bringing into play all different techniques in the same project evokes natural articulation in everyday settings, thereby collecting authentic speech and minimizing the obstructive effects of the *Observer's Paradox* (Labov 1972a:61). The term *Observer's Paradox* denotes the researcher effect, context effect or observer effect, which may greatly lessen the efficiency of data collection. Labov (1972:209) stressed that 'our goal is to observe the way people use language when they are not observed, yet we can obtain these data by systematic observation'. Labov warns that normally, systematic observation of speakers always implies a formal context which provides them with an opportunity to monitor their speech more than in observed situations. The solution to the paradox is to conceptualize the style shifting on a single continuum that ranges from careful to casual or informal styles which informants may tend to avoid in talking to an outsider such as a fieldworker. The continuum is frequently referred to as 'Attention-to-speech continuum' (Labov 1972:211). This study accordingly employed several techniques to capture a representative sample of data.

3.3.1 Sociolinguistic Interviews

In sociolinguistic research, Labov used various techniques, which were later refined, for eliciting naturalistic data on language in daily use. The most influential method ever used was the sociolinguistic interview;; whereby a set of questions was used to elicit as much free conversation as possible. Sociolinguistic interviews involve open-ended questions which are intended to reduce the distance between the interviewer and the subject, making the interaction more natural. Recorded sociolinguistic interviews offer several advantages. The noticeable advantage is their permanency, as it is possible to record and store the data on a computer if one is using appropriate digital equipment, and to re-listen and re-use whenever the information needs to be reaccessed. Recording interviews is less prone to interpreting filtering than taking field notes, which can become subjective. This opinion favours what has been illustrated by Labov (2005:2-3) when he refers to the *Principle of Accountability*, i.e. the principle that "reports of the occurrences of a variant must be accompanied by reports of all non-occurrences and both of them should be accounted for". So successive quantitative analyses are often necessary to elicit whether a particular variant occurs in a systematic pattern or not, and to what degree (Feagin 2002:26).

During the process of data collection the "Labovian" model of the sociolinguistic interview was incorporated. Labov prepared questions which elicited as free conversation as possible. To overcome the unnaturalness in language, Labov (1972:209) advocates the diversion of an interviewee's attention from speech in order to elicit his true vernacular. In support of his argument, his subjects were requested to talk about street games or narrate life-threatening experiences. Questions like "have you ever been in a situation where you nearly lost your life? Where you thought 'This is it!'" involve the subject in a different context and recreate strong emotions which allow the vernacular to emerge.

In the current study interviews were conducted with 56 informants (one informant was interrupted by family commitments right after reading the wordlist and was not available for interview). The interviews took place in the homes of the participants. Most

of the conversations with peasants took place in their homes, especially during the evening hours, to allow them time for their farming activities in the morning. To make contact with the pastoralists, who live away from other people in mountainous bushes, the interviewer with the assistance of her father spent several hours very early in the morning visiting the pastoralists before they went out for animal grazing. In a few cases participants like businessmen were interviewed in their place of work, like shops, restaurants etc. The professionals were targeted during their weekends. The interviewer visited the participants, and the interview took place instantly if the subject had available time, otherwise an appointment was requested.

Before the interview, the prospective informants were informed that they would be recorded and that the interview questions would focus on their life experiences and their daily activities. They were told that the intention was to study the language which would finally lead to documenting it.

To get informants started I decided to use the wordlists first and proceeded afterwards with the background questions of the interview. Questions were intended to give the informants a chance to settle into the idea of the interview before the extended narratives were given. Sometimes the conversations were extended a bit to allow excited interviewees a chance to wind down and calm down. The conversations started with the informants being requested to give their names, their birth dates and the kind of economic activities they do. Feagin (2002:30-31) stresses that the researcher needs to familiarize himself with local customs in order to develop questions which are adjusted to the local conditions. The researcher, being an insider, knows the traditions and the customs of the community, including what not to talk about as far as morals, tradition and customs are concerned. She also knows what is sensitive information with certain age groups or genders is as it could annoy some informants if they were asked to disclose their innermost personal details, or the researcher could be taken as being not respectful or reliable. Questions on marriage issues would be offensive if they involved unmarried females above 25 years, as it is taken for granted that it is unusual to find unmarried women of this age in the society. Those who are above 60 years, and are still in touch

with the traditions of the community would not like to disclose the number of their children, as counting someone's children is taken as a taboo, and believed to lead to death in the family. The researcher was accordingly cautiously selective about what information to elicit unless the informants were ready to disclose deeply personal things such as the death of a husband, a child or a parent. Generally, the informants were requested to give details of their background, talk about their families, their daily activities, their schooling, their marriage ceremonies, entertainment and what they cherished most in their life. They were also asked to narrate stories about their ancestors, festivals, and personal events that they would never forget. In this session informants, like businessmen, revealed stories about car accidents, injuries and robberies they had encountered when they were travelling. Young informants talked about entertainments like watching and playing football, including national and famous international football teams they have watched; movies and music, or playing pool. Farmers revealed stories about natural catastrophes such as floods and the coming of locusts and drought which hindered their successes in farming activities. Pastoralists talked about wild animals like leopards, hyenas and various kinds of snakes which threatened them and dragged off their animals in the bushes. Some of the pastoralists' stories were the same; especially those ones which were abnormal, like the coming of the elephants, which do not harm their animals, and the coming of wild dogs, which took goats from helpless pastoralists. These kinds of stories were associated with superstitions. They evoked however, the feelings of the narrators and made them switch to less formal and natural conversational and narrative style.

It is intended that, in sociolinguistic interviews, narratives are elicited from the subjects on a specific set of topics in order to maximally evoke unself-conscious speech while controlling topic choice across conversations. However, the interviewer did not interrupt the participants when they occasionally strayed away from the interview topic to issues that held greater interest for them, as long as the goal of eliciting natural speech was attained.

In general the recorded sociolinguistic interview proved to be a useful technique in my research. However, since the data was collected in a rural community, certain

impediments were encountered. The main setback was that, some informants were reluctant to be recorded, and associated the recording with superstitions concerning taking someone's voice and his image (known as *kutika kisuka*, meaning "making someone a ghost"*).* In this study, questionnaires lasted for an average of 8 to 10 minutes, the word list 2 minutes, while free conversation was conducted for a minimum of 10 minutes. Free conversations were short according to Labovian standards, as it was truly difficult to get people to speak for longer than 10 minutes given the taboo around tape recording. Though all the participants were native speakers of Chasu, some intentionally used Swahili even when they were requested to use Chasu during the conversation. This resulted in having a number of variables from Swahili which had to be considered irrelevant for the study of phonological variation, though they served other purposes, particularly the study of the effects of language contact. As the interview sessions were the first to be conducted in the community, participants took it as a strange and novel experience. The interview questions sounded pointless to some of the participants, as they were not directly connected with language, but rather inquiring about the details of the life of individuals.

3.3.2 Word Lists

As mentioned earlier, use of a word list is a technique that can be employed when one is dealing with acoustic phonetic investigation (Feagin 2002:31). This involves presenting subjects with lists of words containing key variables. Taking into consideration that part of this study is phonologically oriented, word lists proved useful in this research. Two lists of 20 words each were used to examine the articulation of phonological variables (s) and (z). When subjects are asked to read lists of words which contain phonological variables that connote sociological stigmatization in the society, they tend to concentrate on their pronunciation to avoid mistakes so as to sound formal. The informants are also likely to monitor their pronunciation when the researcher discloses to them that the reading list of words includes certain variables which indicate the way the informants articulate. Therefore, in order to minimize speech monitoring, 15 words in each list contained the phonological variables to be examined, while 5 words

were ‘distracters’ from each list having nothing to do with the sounds (z) and (s). All subjects who read the words from the list were recorded using an MP3 digital recorder, sometimes accompanied by a Camcorder. The recorded data were finally downloaded to the computer for storage and transcription.

This technique was very useful as all the informants had to read the same list of words, which led to easy quantification of data. However the possible drawback of word lists is the formality and self-consciousness they sometimes create, which lead to possible unnaturalness of the articulations. At the same time some informants, especially pastoralists and peasants, were unable to read, due to illiteracy or poor sight because of age. Three informants who could not read were involved in sociolinguistic interviews only. Some informants claimed to be able to read but could not, or read poorly. In order to avoid embarrasing them, those subjects who were willing were requested to repeat words after the researcher. The word list requires patience and careful attention in a rural society where illiteracy is evident.

3.3.3 Participant Observation

This technique of data gathering involves making the researcher a member of the group he is observing so that he can share some of their experiences (Kothari 1990). Participant observation normally serves as a complementary method of data collection as well as being one of the techniques which evolved for reducing formality in face-to-face interviews. A technique that evolved from anthropological linguistic studies, participant observation requires analysts to integrate themselves within the community under investigation either by engagement in local affairs and /or developing personal associations with members (Feagin 2002:33). Concerning participant observation, Dewalt and Dewalt (2002:1) give the following definition:

> For anthropologists and social scientists, participant observation is a method in which a researcher takes part in the daily activities, rituals, interactions, and events of a group of people as one of the means of learning the explicit and tacit aspects of their life routines and their culture.

For linguistics, however, the main emphasis falls on observing and noting down linguistic behaviour in context while behaving like a member of the community. It would be difficult for an outsider to become a participant observer, unless they were fluent in the language of the speech community.

In order to supplement the data gathered through interviews and wordlist techniques, participant observation was also employed in this study. This was to observe the language use in public settings, with some recordings of free conversations. This involved attending social gatherings like funerals, local markets which take place every Wednesday, shops, religious meetings etc. Of all the participatory observations I made, those gathered at the local weddings, attending to the flour milling machine and paying visits to friends and relatives, were the most valuable. I spent ample time, cooking, cutting firewood, fetching water, eating and chatting with people. Hence I managed to record some of the conversations, and note down some words with variable sounds. In events like local weddings I had to request permission from the holders of the event to attend and inform them of what I was doing. Up to 10 intended observations, four of which were recorded, were made to close up any gaps in the one-on-one interviews and wordlists. Apart from observations in shops, markets and religious meetings, where people mingle to some extent, other participatory observations were gender biased. This is because when performing local and traditional ceremonies such as weddings and funerals, customarily women do not always sit together with men, though all attend those ceremonies. So the researcher, being a woman, had more time with females except in shops and markets.

One of the prime advantages of participant observation is that the researcher becomes something of an insider of the group under study. Consequently, the effects of the *Observer's Paradox* are greatly lessened as the *Observer* has almost vanished, and

authentic contexts already exist in which speakers will produce authentic language. Notwithstanding such optimistic facets, there is need for qualification of this methodology. In as much as note-taking is restricted, the researcher is not able to note every token of interest during the course of the observation process. Secondly, writing down linguistic examples (sentence, word, or phrases) during participatory observation nevertheless entails a degree of selectivity and hence subjectivity. Feagin (2002:34) notes that this could affect objective data quantification. Furthermore, the researcher might not always be able to note down the background details of the speakers, which also leads to lack of meaningful sense to the quantified data.

However, this technique proved useful in my research, as it presented an opportunity to interact with members of the communities studied and hence I closely observed their cultural practices and interaction with different groups in real life contexts.

3.3.4 The Questionnaires

This is a method of data collection in which a number of questions are typed and printed in a definite order on a form or set of forms. The researcher can administer the questionnaires herself or mail them to the respondents. The questionnaire may consist of open ended or closed questions. For easy administration and inexpensive analysis, multiple choice or *Yes* or *No* questions are used to obtain specific information.

In this study the questionnaire method was used to gather personal information concerning the respondents and their language use. Ten questions were administered, 8 multiple choice and 2 informational. The questionnaire is attached as appendix 1. In questions 1, 2 and 5, the respondents were to indicate their sex/gender, age group and level of education. Questions 3 and 4 were about their ethnicity/ clan and occupation while questions 6-10 were about language use. The respondents had to state which of the languages they spoke before going to school, which one they used at home, in the work place, and with friends and with relatives. The languages of choice were Chasu, Swahili, English and combinations of two or three of those. The first five questions would help to

categorise the respondents into the social groups they belong, while the rest would assist in examining the question of the effects of language contact in this ongoing shift to becoming a multilingual society.

The questionnaire was administered before the actual interview and the reading of the wordlist. The questions about the background of the informants were used to warm people up and get them to relax. The informants also responded to questions about their occupation, thus triggering the questions to ask them during the conversation.

This method helps to give general information about the interviwees such as age, gender, education levels, occupation, etc. Its importance is also to ensure the inclusion of a variety of people in order to get a genuine representation.

My final sample constituted 57 informants who participated in questionnaires and wordlists, of which 56 participated in free conversation, while 10 participatory observations were made, of which 4 were recorded.

3.4 Data analysis

After the interviews, data analysis started with systematizing the corpus and giving labels to each interview and pseudonyms to interviewees. This was followed by the process of listening to the interviews and noting down passages for transcription. The conversations were written down from the recorded text using Chasu orthography. This orthography is based on the existing Chasu literatures like Bibles and hymn books. Southern Chasu orthography is the same as the one used in Swahili; however, certain sounds from data were transcribed using the IPA system. In Chasu, there is no orthographic character for the sound /ž/ (IPA [ʒ]); so where it occurred in the data (ž) was used instead. Since this study has two main objectives, the analysis was divided into two main sections. The first section is concerned with language use and the variation in the context of language contact in the Southern Chasu dialect, while the second is concerned with phonological variation.

3.4.1 Data analysis for variation in the context of language contact:

After the transcription of the data, borrowed lexical items were identified from both Swahili and English. Strings of code-switching between Swahili and Chasu or with English were also identified, followed by an exercise of coding these items into different factors of age groups, social class, education difference and sex. The questionnaires about language use were also statistically summarized in percentage. The results and the discussion are presented in chapter four.

3.4.2 Data analysis for the phonological variation

The transcription for the phonological variables started during the time of data collection, in order to ascertain whether the conversations adequately captured the variables intended. As stated in chapter two, VARBRUL was used to analyse the data for this study. The first step when carrying out the VARBRUL analysis was to identify the variables. After the transcription procedure, all Southern Chasu words with the variable sounds (s) and (z) were identified. The variants were also identified as follows: (a) (s) was realised either as a voiceless alveolar fricative [s] or voiceless dental fricative [θ], (b) (z) occurred as either a voiced alveolar fricative [z] or voiced dental fricative [ð]. The application value for the variants were; (0) represented [s] while (1) was for [θ] for the first run and (0) represented [z] and (1) for [ð] for the second run. The next step was to specify the factors which may potentially influence the choice of the variants, i.e. the independent variables. These included both external (social) and internal (linguistic) variables.

(a) External variables and the hypothesis

Gender: In Western industrialised countries gender/sex has been used to examine the variation of the phonological variables. In most sociolinguistic studies as reviewed in chapter two, women use the linguistic features closest to the standard language more frequently than men of the same social class, age and education level (Trudgill 1983:161). Women are considered to trigger language change compared to men. In this rural sociolinguistic analysis, it is hypothesized that, given the same social factors, men use standard linguistic variants more frequently than women.

Age groups: Speakers were categorized into three main groups of age difference namely young: 21-40; middle age: 41-60; and old age: 61+. Though old and middle-aged speakers may activate language change in the society, in this analysis it is anticipated that younger speakers begin to adopt more standard variants and hence prompt language change.

Social class: Based on their occupation types as stipulated earlier, speakers were categorised into four main social economic groups, including professionals (teachers, nurses, leaders etc), business people, peasants and pastoralists. In this analysis, it is predicted that professionals and business individuals make greater use of standard features [s] and [z] while pastoralists and peasants use the non-standard [θ] and [ð] more frequently.

Education levels: Four main groups were formed, from tertiary education, high secondary school level, primary school, and those who did not go to school at all. In this analysis, the tertiary level of education and the high secondary school speakers are more likely to use standard variants [s] and [z]. The non-standard variants [θ] and [ð] are expected to predominate among speakers who have primary school education only, and those who did not school at all.

Language use status: As stipulated in chapter one, this community is mainly bilingual, with a few monolingual individuals. Hence in this analysis we essentially categorised two

groups: monolinguals who use Chasu only, and bilinguals speaking Chasu and Swahili. We anticipate in this analysis that bilinguals use more of [s] and [z] while the monolinguals use [θ] and [ð] respectively, as it is assumed that [s] and [z] are Swahili while [θ] and [ð] are indigenous.

Style: In this analysis, we have two distinct styles of language use, wordlist style and casual style (based on conversation). As expected, when speakers read the wordlist or passages (or repeat words after the interviewer) they tend to sound formal, while in conversation they produce more natural speech. Therefore, it is hypothesized that in casual style, speakers use more of the non-standard variants, while in wordlist syle they use the formal standard variants.

(b) Internal variables

In this analysis we also included internal variables such as (a) the occurrence of the variants in word structure (whether initial syllable, medial or final syllable), (b) vowel after the variables (back or front vowels) and (c) status of lexical item (whether from Chasu or borrowed words from Swahili). Chapter five presents and discusses the results of the quantitative analysis of data concerning the phonological variation.

CHAPTER FOUR
VARIATION IN THE CONTEXT OF LANGUAGE CONTACT IN SAME DISTRICT

4.0 Introduction

In chapters one and two, the relation between Swahili and the ECLs was outlined, as well as how Swahili is associated with power, modernity and dynamic life. This chapter therefore, examines how social factors operate as regulators of language variation in the context of language contact in the trilingual Chasu society. It includes three main sections. Firstly, it describes the predominant language choices in the rural Chasu community. Secondly, it investigates the prevalence and frequency of occurrence of lexical borrowing and code-switching as language contact phenomena. Thirdly, it examines to what extent social factors such as social class, education levels, age difference, and gender function in relation to lexical borrowing and code-switching.

4.1 Language choice in rural Chasu community

The question of language choice at the societal level arises where two or more distinct languages or varieties co-exist. Fishman (1972) identifies determinants of language choice to be (1) the social group to which an individual belongs, (2) the context or setting at the moment of conversation, and (3) the topic of discussion. These determinants are connected to language domains, which may differ from one speech community to another. Examples of domains are the home, work, school, religious and ritual ceremonies, festival and cultural events etc. In many communities, additional factors related more to characteristics of the individual rather than of the community are operatives, like age or the identity of the interlocutor (Dorian 1981:80).

Concerning language choice, the data in this section was obtained through questionnaires. Out of 10 administered questions, questions 6-10 were based on the domains related to language use. The informants were asked to state which of the

languages they spoke before going to school, at home with family members, in the work place, with friends and with relatives. The languages of choice were Chasu, Swahili, English and combinations of these. The distribution of code choices is as presented in table 4.1 below.

Table 4. 1: Language choice in contemporary Chasu community by number of speakers (n = 57)

Domains/languages	Chasu	Chasu and/or Swahili	Chasu, Swahili and English
Speakers per language	3 (5%)	40 (70%	14 (25%)
Language used at home	34 (59%)	14 (25%)	9 (16%)
Language used at work	26 (46%)	19 (33%)	12 (21%)
Language used with close relatives	41 (72%)	16 (28%)	0
Language used with friends	13 (23%)	32 (56%)	12 (21%)

In this small survey there were 3 monolinguals, who did not have formal education at all and who worked as pastoralists. Another 40 informants were bilinguals in Swahili and Chasu, working mainly as subsistence farmers or in small businesses. There were 14 informants with English, Swahili and Chasu in their language inventory, of whom 13 were working in professional activities and 1 was a businessman. Currently, numerous people, mainly young speakers (male and female) are bilinguals in Swahili and Chasu, while only a few are Chasu monolinguals as table 4.2 illustrates.

Table 4. 2: Language status by age groups and sex

Age groups	Sex	Chasu	Swahili and Chasu	Chasu Swahili and English
Young (21-40)	F		9	3
	M		8	3
Middle (41-60)	F		8	2
	M		6	5
Old (61+)	F	1	3	-
	M	2	6	1

Only 2 male and 1 female speakers aged above 61+ years are monolinguals of Chasu. A combination of young, middle-aged and one old-age speaker are trilinguals in Swahili, English and Chasu. The total number of female trilingual speakers is roughly the same as that of male, in the young and middle-aged groups. Eight informants claimed not to use any other language except Chasu, despite being bilinguals in Chasu and Swahili via schooling.

Concerning language use at home, it has been explained in chapter one that ECLs in Tanzania are reserved for family discussions and traditional functions such as marriage ceremonies and rituals in the villages, where family members are highly involved (Rubanza 1979). In urban Kenya, Myers-Scotton (1993a:39) demonstrates that workers code-switch when they are with fellow workers. When they return home to join their family members, most Africans use an ECL almost exclusively, unless their marriage is inter-ethnic or they are highly educated. Those at the top of the economic scale might speak English with their children. In this study, table 4.1 illustrates that 34 informants use Chasu only, 14 use both Swahili and Chasu, while 9 use Swahili, Chasu and English for their home conversations. Currently, there is no extensive intermarriage (with other ethnic groups) in rural Chasu community, which could trigger a combination of

languages, or the predominance of Swahili as the intermediate language among family members. Recently, in Tanzania, there has been a rapid increase in English-based private schools from kindergarten through primary to secondary levels, spreading from urban areas to rural villages. This goes hand-in-hand with improved transport systems, as well as telecommunication schemes. Peasants and pastoralists use Chasu in their homes and send their children to the ordinary primary schools. This does not imply that peasants are not competent in Swahili, but the fact is that they do not practise it since there is no motivation to do so. Individuals who run businesses use Swahili most of the time, even if their highest level of education is that of primary school. They regularly combine Swahili and Chasu due to their daily external business attachments. This combination is carried into their home. During the interviews, some highly educated informants claimed that they placed their children in an English-medium school and used English at home with their children. In this kind of family, Chasu is used only when grandparents pay a visit to the household.

Concerning language used at work, table 4.1 indicates that 26 informants used Chasu, 19 used both Swahili and Chasu, while 12 speakers used a combination of Chasu, Swahili and English. This distribution reflects the nature of the economic activities among the informants. Teachers use English and Swahili in schools where these languages are the predominant media of instruction. Peasants and pastoralists consistently use Chasu, except during transactions involving their farm products with people from outside the community. Businessmen and women use a combination of Swahili and Chasu depending on the types of customers involved.

Pertaining to the language used with relatives, Chasu is reserved for this domain. Forty-one informants expressed their preference for Chasu in conversations with relatives, while 16 favoured both Swahili and Chasu. Observation shows that though family members would wish to maintain Chasu with relatives when sensitive family issues are discussed, there are certain families which have relatives based in the urban areas, where, as noted previously, Swahili is dominant. These relatives sometimes return

to the village with their children who are not able to converse in Chasu. This increases the use of Swahili among relatives.

Table 4.1 also indicates that 13 informants used Chasu in conversation with friends, 32 informants used Chasu and Swahili, while a combination of Swahili, Chasu and English was used by 12 informants. This is a reflection of people with whom individuals make contact in daily informal conversations. As professionals are likely to make friends with their fellow elites, the use of Swahili and English is inevitable. On the other hand, businessmen and women use Swahili and Chasu with friends whom they meet daily through their business activities. A few locally-based people such as peasants and pastoralists retain Chasu with friends.

This section has presented a brief description of a distribution of language choice in the Chasu community. Language choice seems to be connected with functions performed by a distinctive social group in particular contexts. It is also linked to socio-economic activities, and educational attainment. Thus Chasu is only used by individuals for certain domains, while a combination of Chasu and Swahili or Chasu, Swahili and English is used in other domains. It is explained in chapter one that language use in different domains results in differential social status. English is socially considered 'High' in relation to Swahili and Swahili is 'High' in relation to Chasu (double overlapping diglossia). Together with this pattern, the data tells that there is a group of people who use Chasu, Swahili and English (a triglossic pattern). Gumperz (1969:2) states that 'each time minority language groups come in contact with majority language groups under conditions of rapid social change' we find code-switching. The main focus in the next section is to examine how contact between Chasu, Swahili and English in conversation results in lexical borrowing and code-switching in the Chasu speech community.

4.2 Lexical borrowing and code-switching

In distinguishing lexical borrowing from code-switching, Poplack and Meechan (1995:200) assert that, 'code-switching may be defined as the juxtaposition of sentences or sentence fragments, each of which is internally consistent with the morphological and syntactic (and optionally, phonological) rules of its lexifier language… borrowing is the adaptation of lexical material to the morphological and syntactic (and usually phonological) patterns of the recipient language'. Consequently, the categorisation of lexical borrowing and code-switching in this study is based on the supposition that code-switched elements occur when a bilingual introduces completely unassimilated forms into the structure of a receiving language, whereas borrowed elements are nativized by assuming the morphological, syntactic and often the phonological identity of the recipient language (Haugen 1956:40, Thomason 2001:134). Despite this assumption, the status of an expression as code-switching (particularly single-lexeme items) or lexical borrowing items is not always straightforward. Its complexity increases when there is an attempt to distinguish Poplack's idea of 'nonce borrowing', sometimes known as 'momentary' borrowing (single-word code-switching) from lexical borrowing. Poplack et al (1988:58) classifies nonce borrowing as 'a one-off occurrence resorted to by the speaker', different from established lexical borrowing which is 'recurrently used in the community'.

Traditionally, the degree of integration of loan forms into the recipient language has been the criterion used to distinguish borrowed forms from single-lexeme switches. Myers-Scotton (1993b:162), however, takes a contrary view. She suggests the frequency of occurrence to be a reliable decisive factor. Lexical borrowed forms are usually recurrent and widespread in the speech of the individuals across the community, including monolingual speakers of the recipient language. Monolinguals access loanwords regularly, along with the remainder of the recipient-language lexicon. Myers-Scotton (1995:41) suggests further that, despite their common origins within the donor language, borrowed forms and code-switching also differ in terms of predictability and recurrence. Lexical borrowing is predictable since it has a status in the lexicon as belonging to the recipient language. Code-switching, on the other hand, has no

predictable value: items involved in code-switching may not necessarily recur. Poplack (1980 & 2004:5-6) affirmed that code-switching is used by those individuals whose language skills in both languages are stable. While lexical borrowings occur only at the level of word category, code-switching extends from the lexical level to that of syntax or utterances. Like code-switching, nonce borrowing is neither recurrent nor widespread and also requires a certain level of bilingual competence. Therefore, since nonce borrowing has certain common attributes with code-switching, they are grouped together in this analysis, while recurrent lexical borrowing is treated as another group in this study. However, to distinguish switches of single-lexeme from borrowed words from Swahili was not straightforward, particularly where certain grammatical features were common to Swahili and Chasu. For instance, the existence of noun class 9/10 in both Chasu and Swahili nouns imposes some difficulties in differentiating the lexical borrowed word from single-word switches; this will be addressed in detail in sub-section 4.2.1.

4.2.1 Lexical borrowing

In situations of language contact, various kinds of items are borrowed, with lexical material the most easily borrowed. Myers-Scotton (2001:41) explains that lexical borrowing occurs when 'L1 speakers of the less prestigious groups take into their language words from the L1 of the more prestigious language'. Though the reverse direction is also possible, Myers-Scotton (2006:209-11) adds that the exchange between languages is never equal; a group of people with less opportunity in socio-economic status or political control are likely to borrow the most.

Together with other traits, borrowed words may be assimilated or partially assimilated into the phonological, morphological and syntactic identity of the recipient language, though with a different degree of integration between grammatical categories. Myers-Scotton (2002:42) comments that the degree of integration depends on a number of factors, including phonological difference between the two languages, but also the degree of bilingualism of the speakers using the borrowed forms. In a now classic study, Haugen (1956:55 & 1972:79) states that, when borrowed items are first adopted, they

tend to have an uncertain linguistic status for some time. Each individual in the community may adapt a particular loan word to a varying degree before it meets with more general social acceptance. Haugen proposes that with phonological adaptation, the main mechanism is the replacement of the donor language phonemes by the phonemes of the base language, in three stages. First, the lexicon is introduced by the bilinguals in a phonetic form which is as close to that of the model as possible. Then the lexicon may be integrated once other speakers start using it, and foreign elements of the word will be substituted by the native ones. A later stage may involve the use of the word by monolinguals. At this stage there is practically complete substitution at the phonetic levels**.**

In this study, as mentioned before, three languages are in contact. Chasu is the recipient language, while Swahili and English the donor languages, at least for new words. Generally, when a word is borrowed from Swahili into Chasu, the phonological effects are minimal except for phonemic tone. Swahili does not possess tone and it is claimed that tone was lost in its earlier stage, probably under the extensive influence of Arabic (Vitale 1981:10). Loanwords are assigned tone when they are borrowed into Chasu. Otherwise, rules for the pronunciation of forms resemble those that of Chasu. In writing, the orthographic forms resemble each other, since tone is not usually marked in Chasu spelling. The sound patterns and syllabic structure of Chasu and Swahili are characteristically Bantu. Both have a CVCV syllable system in which it is easy to insert words or syllables from one language into another. Some words are likely to retain their phonemic structure while others adopt new inflectional affixes. Morphological and syntactic adaptation does occur in loanwords from Swahili into Chasu, as I will show in table 4.10 below. On the other hand, when Chasu borrows words from English, some assimilation is necessary, since the phonemic systems of these languages are different. Examples of borrowed forms are in table 4.3 below:

Table 4. 3: English borrowings into Chasu and Swahili

SN	Chasu	Swahili	English
a.	thieta	chumba cha upasuaji	theatre/operating room
b.	presheni	upasuaji	operation
c.	timu	timu	team
d.	kocha	kocha	coach
e.	hospitali	hospitali	hospital
f.	kozi	kozi	course
g.	m-kristo	m-kristo	Christian
h.	blauzi	blauzi	blouse
i.	i-shati	shati	shirt
j.	sketi	sketi	skirt
k.	sekondari	sekondari	secondary
l.	tarumbeta	tarumbeta	trumpet
m.	kwaya	kwaya	choir
n.	va-jerumani	wa-jerumani	Germans
o	i-begi	begi	bag/luggage
p	picha	picha	picture

In table 4.3 the borrowed lexicons from English, *thieta* in (a), *presheni* in (b), *timu* in (c) and *kocha* in (d) have been adapted to the Chasu phonemic system. Chasu and Swahili observe the open-syllable rule common in Bantu languages. Thus there is an insertion of epenthetic vowels where there are consonant clusters in English borrowings, as in example (l) *tarumbeta*, (n) *va-jerumani* and (p) *picha*. In addition, a word-final vowel is inserted where necessary. This happens to all grammatical categories.

Sometimes English words may pass via Swahili into Chasu, resulting in very similar loan forms in Chasu and Swahili, as in borrowed words *kozi* in example (e), *i-shati* in (i), *tarumbeta* in (l) *kwaya* in (m) and *sekondari* in (k). These have adapted the syllabic system for both Chasu and Swahili. However, a few changes may occur when prefixes for plural and singular are inflected into the lexicon, like for *i-shati* in (i) *va-.jerumani* in (n) and *i-begi* in (o). Otherwise the lexicons remain the same in both Chasu and Swahili.

Sankoff et al (1990:76) suggest that established borrowings may retain some traces of the phonological patterns of their language of origin. In this study, there are a few cases where a string of consonants from the English syllabic system is partially assimilated to Chasu or Swahili. The open-syllable rule of Chasu is not observed, as vowels are sometimes not inserted in consonant clusters of English borrowings. Thus some English consonant clusters are maintained in borrowed words like *hospitali* in (f), *m-kristo* in (g), *sketi* in (j) and *blauzi* in (h). Since not all borrowed lexemes from English receive a full integration into Chasu, we can predict the co-existence of different phonemic sub-systems in Chasu. However, the vowel is always inserted in the syllable of the word final position. The morphological and syntactic assimilations which frequently occur with words from Swahili will be discussed below, where I examine why some grammatical categories are more easily borrowed than others.

In this study I have classified two main categories of lexical borrowing based on the terms 'core' and 'non-core' vocabulary, used by Appel and Muysken (1987:165) or their equivalent; 'core' and 'cultural' borrowing (Myers-Scotton 2002:41, 2006:215-217) or 'basic' and 'non-basic' (Thomason 2001:68-74). Core borrowings are words that 'duplicate elements that the recipient language already has in its store' (Myers-Scotton 2006:215), i.e. concepts already existing in the recipient language. In situations of unequal prestige, it is the borrowing language which is usually lower in status. Consequently, core borrowings are likely to be used for reasons of prestige prompted by a desire to identify an individual with the donor language and its higher social status. In comparing core borrowings with code-switching, Myers-Scotton (1993b:169-76) asserts that in their initial stage these forms are identical. They may appear once or twice in a large data corpus with no predictability as to their recurrence. They are mainly used by speakers with fluency in both languages. Myers-Scotton (1993b:174-75) hypothesizes that, before they became established borrowings, such core lexemes were code-switching forms with connotations of prestige. They achieve the status of loan words by recurring over time in the speech of more and more individuals. Not all code-switched forms become core borrowings, only the recurring and assimilated ones.

Cultural or non-basic borrowings are words which make reference to new concepts and objects to the culture. Myers-Scotton (2002:41) asserts that core borrowings enter the recipient language gradually through code-switching, commonly in bilingual speech. On the other hand, cultural borrowings appear relatively quickly, when influential people start to use them. They are then rapidly integrated into the recipient language and spread to both monolinguals' and bilinguals' speech. In this study, non-core borrowings can be divided into certain semantic fields; for example, entertainment, geographical and medical terms. One of these patterns is the example 1 below, on the topic of football. The conversation is with a 34-year old businessman, who also works as a conductor of a truck which travels through different provinces in Tanzania. In data presentation in this chapter; borrowed words and switched forms from Swahili are labelled in *italics*, borrowed words or switched forms from English are in **bold** form, while the English translations are given in 'single quotation marks.**'**

Example 1

Q: Ni kintu ani wekunndie kuronga zaidi ya ivi vintu vyekunka kipato?
'What do you like to do beside your activities which earn you'?

A: Iki aha mi ni *mchezaji* wa mpira na nina na **timu** yangu ya Miamba star mpaka iki aha tuho he *ligi* na Jumapili tu-a-*cheza* **fainali.**
'Currently, I am a football player, and I have my team, Myamba Stars, meanwhile we have a League, we will play the Final on Sunday'.

Q: Hena kipindi wa-na-*cheza* ukafunga **magoli**
'Has there ever been a time when you played and score some goals'?

A: Mi nafasi yangu ni **beki**.
'My position is at the back'.

Q: We u **beki**?

‘Are you a defender? (back position)’

A: Kwa maana ni-*cheza* nafasi ya vurinji.
‘… meaning that I play a defence position’.

Q: Urerehia *mipira*?
‘Do you watch football?’

A: *Mipira* nirerehia sana nekifika Dar es Salaama nirerehia nekifika hata Morogoro, Arusha... vilevile nire mpenzi wa *luninga* sana kurerehia *mipira* ya Uingereza **ligi** mbalimbali ža Uingereza na *makombe a Dunia.*

‘I watch football when I arrive in Dar es Salaam, Morogoro and Arusha… I watch I like watching through TV, United Kingdom’s football matches and League as well as World Cup’.

Q: U shabiki wa **timu** ani?
‘Which team are you supporting?’

A: Ushabiki wangu kwa iki aha he *i-taifa* letu ni *shabiki* wa Simba
National-wise, my support is currently for Simba’.

Q: **Timu** ya nze?
‘Which outside team’?

A: **Timu** ya nze... mi niho Manchester.
‘Outside team... I am with Manchester’.

From this chunk of conversation we find a set of lexical items associated with football. These include **timu** ‘team’, **ligi** ‘league’, **beki** ‘back’, **ma-goli**, ‘goals/scores’, **fainali** ‘final’, which are borrowed from English via Swahili . Words like *cheza* ‘play

football', *luninga* **'**television' and *shabiki* supporter/fan', *ma-kombe a Dunia* 'world cups', *mpira* 'football' are directly borrowed from Swahili.

The extent to which a speaker borrows either core words or non-core words from another language is directly proportional to the intensity of contact between the recipient and the donor language. Thomason (2001:70-74) clarifies how frequency of contact relates to borrowing scales by outlining four stages of contact. The first involves casual contact: borrowers do not need to be fluent in the donor language and only non-basic vocabulary, most often nouns, are borrowed. A second stage is that of slightly more intense contact, where the borrowers must be bilinguals but are likely to be a minority among the borrowing-language speakers. Here function words and a few non-basic content words are borrowed. A third stage involves more intense contact. It is characterized by more bilinguals with positive attitudes to and other social factors favouring borrowing. Basic words, such as nouns, verbs and adjectives, which may be present in both languages, may be borrowed at this stage. The fourth last stage involves intense contact, with very extensive bilingualism among borrowing-language speakers, with social factors strongly favouring borrowing. Heavy borrowing is continued in both lexicon and structure. Though each bilingual context has its own peculiarities, generally it is content words that are borrowed, while function words are more resistant. Appel and Muysken (1987:171) add that the mass of established loan words consist primarily of nouns, verbs and adjectives, with a number of adverbs since they are linked to content words, while pronouns are hardly ever borrowed nor articles, quantifiers, demonstratives and prepositions.

Table 4. 4: Lexical borrowings from English and Swahili into Chasu

Lexical category	Core borrowings	%	Non-core borrowings	%	Grand total	%
Nouns	236	57.8	584	91.1	820	78
Verbs	140	34.3	57	8.9	197	18
Pronouns	-	-	-	-	-	-
Adverbs	20	4.9	-	-	20	1.9
Adjectives	6	1.4	-	-	6	0.57
Conjunctions	6	1.4	-	-	6	0.57
Interjection	-	-	-	-	-	-
Total	**408**		**641**		**1049**	

In this study there was a total of 1049 loanwords, used by bilingual and trilingual Chasu speakers, of which 127 were English words borrowed via Swahili and 924 were direct from Swahili. Table 4.4 above shows the distribution of 1049 loans;, 641 of these were non-core borrowings, and all are content words, made up of 584 (91%) nouns and 57 (8.9%) verbs. There was a total number of 408 core borrowings of which 236 (57.8%) were nouns, 140 (34.3%) verbs, 20 (4.9%) adverbs, 6 (1.4%) adjectives and 6 (1.4%) conjunctions.

The Chasu data shows that the content words, i.e. verbs and nouns, from Swahili and English are borrowed more frequently than functional words both in non-core and core borrowings. Table 4.5 below shows the most frequent core and non-core loanwords from Swahili to Chasu, and that non-core lexemes, particularly nouns, occur most frequently. It is worth mentioning that tables 4.4 above and 4.5 below include all tokens, including the repetition of the same item.

Table 4. 5: The frequency of occurrence between core and non-core borrowings

Core borrowings	Gloss	Frequency	Non-core borrowings	Gloss	Frequency
va-zazi	parents	14	shule	school	132
i-kanisa	church	19	i-darasa	class	45
uwezo	capacity	12	biashara	business	34
ma-shamba	farms	8	sekondari	secondary	22
ma-tatizo	problems	12	tangawizi	ginger	20
nyumba	house	7	msingi	primary	16
m-kulima	farmer	6	elimu	education	15
harusi	wedding ceremony	5	mw-alimu	teacher	14
i-tukio	event	11	tarehe	date	13
changamoto	challenge	12	i-gari	a car	13
kazi	work	3	timu	team	7
mawazo	thoughts	3	pesa	money	8
kipato	income	4	thieta	theatre	8

In this table and in the analysis generally, the ultimate origin is not relevant; rather, the word is first adopted in Swahili and later in Chasu. There are words like *mw-alimu, tarehe* and *elimu* which are from Arabic via Swahili, *pesa* and *gari* from Hindi, while *timu, sekondari, thieta* for this context are from English. Table 4.5 also indicates that, while the most frequently occurring core borrowings reflect general topics, non-core borrowings reflect specific topics or certain fields such as formal education, business, medicine etc. These topics are connected to aspects of formal borrowed culture. In this study, specific place names were not counted.

It is accepted in contact linguistics that the large number of noun borrowings might not in fact be a high proportion, but rather a reflection that nouns are the most common category in most languages. This can only be measured through appraising the proportion between loanword tokens and the total number of word-tokens in the recipient language. Table 4.6 is extracted from conversations of five respondents who had a small number of incidences of lexical borrowing and code-switching from Swahili and English. These 5 speakers were chosen based on their extensive use of Chasu in their conversations. The frequency of the lexical category in their Chasu was used to ascertain

the proportions of borrowings beyond raw numbers. A total number of 1401 conservative words from different grammatical categories was analysed from more or less monolingual speech. Of these, 451 were nouns and 564 were verbs, while the rest were the functional word categories.

Table 4. 6: Occurrence of lexical categories in Chasu in a sample of speakers.

Speaker	Nouns	Verbs	Adverbs	Adjectives	Pronouns	Conjunctions	Interjections
1	53	122	8	27	10	21	-
2	134	176	18	34	31	19	1
3	174	176	34	18	75	23	1
4	57	58	7	8	11	6	2
5	33	32	14	7	2	6	3
Total	**451**	**564**	**81**	**94**	**129**	**75**	**7**
Grand Total				**1401**			
%	32.1	40.2	5.7	6.7	9.2	5.3	0.49

In this fragment of the Chasu corpus in table 4.6 above, instances of verbs and nouns are more frequent than other word categories. This implies that in regular Chasu conversation, verbs and nouns occur more frequently in comparison to adjectives, pronouns, adverbs and interjections. It is now possible to compare the frequency of lexical borrowings and word categories to help us to ascertain which word category is most borrowed in Chasu. To facilitate the comparison, the ratio of row 1 borrowings to row 2 (non-borrowings) in table 4.7 was expressed as an index. This was done by dividing the figures in row 1 by their corresponding figures in row 2 and multiplying by 100.

Table 4. 7: Borrowing index in Chasu (for both core and non-core borrowings)

	Nouns	Verbs	Adverbs	Adjectives	Conjunctions
Borrowing %	78	18	1.9	0.57	0.57
Chasu corpus %	32	40	9	7	5.3
Index	243	45	32	8	11

Table 4.7 shows that, though verbs occur more frequently than nouns in Chasu, nouns from Swahili are borrowed more frequently than verbs. In explaining why nouns are frequently borrowed Poplack, Sankoff and Miller (1988) propose that nouns are forms with the most lexical content. Appel and Muysken (1987:171) have argued before that the most important reason for borrowing is to extend the referential function of a language. Thus nouns are easily borrowed, since the referential function of the language is made primarily through them. They also add that content words (nouns, adjectives and verbs) are more susceptible to being borrowed because they have a clear link to the cultural content in comparison to functional words (pronouns, adverbs, conjunctions and interjections).

On the other hand, particularly in the study of Chasu, borrowing may well be enhanced by the effortless way in which certain lexemes can be adopted and integrated into the recipient language phonologically, morphologically and syntactically. Swahili nouns are marked by gender and number in a characteristically Bantu prefix system. Gender is grammatical and affixes mark a noun for membership in a noun class, while number is either plural or singular (Vitale 1981:13). These Bantu traits apply to Chasu nouns as well. Thus the noun class systems of Chasu and Swahili are closely related; hence the borrowed nouns are easily integrated into the recipient discourse. Swahili has the following noun class system in comparison to Chasu.

Table 4. 8: The noun class system of Swahili

Class	Noun Prefix	Examples in singular/plural	Subject prefix in singular/plural	English
1/2	M/Wa	M-tu/wa-tu	Yu-|a-/Wa-	person (s)
3/4	M/Mi	M-ti/mi-ti	U-/I-	tree (s)
5/6	Ji|ɸ/Ma	Ji-na/ma-jina	Li-/Ya-	name (s)
7/8	Ki/Vi	Ki-atu/vi-atu	Ki-/Vi-	shoe (s)
9/10	N/N	Ndizi/ndizi	I-/Zi-	banana (s)
11	U	U-bao	U	blackboard
14	U	U-ongozi	U-	leadership
15	Ku	Ku-imba	Ku-	singing
16	Pa	Hapa	Pa-	place/definite
17	Mu	Humu	Mu-	place/inside
18	Ku	Huku	Ku-	indefinite place

Source: Mohamed (2001:40-51)

In the Swahili noun class system, as shown in table 4.8, noun class 6- is the prefix for the plural forms of 5, 14 and a few nouns in 11, while noun class 10 is the plural form for 9 and a few nouns in 11.

Table 4. 9: The noun class system of Chasu

Noun class	Noun prefix	Examples in plural and singular	Subject prefix	English
1/2	M-/Mw-/Va	M-fumwa /Va-fumwa	A-/Va-	King (s)
3/4	M/Mi	M-ti/Mi-ti	U-/I-	Tree (s)
5/6	I/Ma	i-hemba/ma-hemba	Li-/A-	Corn (s)
7/8	Ki-/Vi-	Ki-ogwe/Viogwe	Ki-Vi-	Potato (s)
9/10	N/N	Nkuku/Nkuku	I-/ Ži-	Chicken (s)
11	Lu-	Lu-kundo	Lu-	Love
12/13	Ka-Vu-	Ka-guro	Ka-Vwa-	Diminutive/dog (s)
14	Vu-	Vu-aži	Vu-	Sickness
15	Ku-	Ku-hema	Ku-	Breathing
16	Ha-	Ha-ntu	Ha-	place

Sources: Kagaya (1989) and Mreta (1998:60)

As shown in table 4.8 and 4.9 noun class 1, 3/4, 6, 7/8, 9/10 and 15 have the same noun class prefixes in both Chasu and Swahili.

Examples of the frequently borrowed lexicon from Swahili and English, and the way they have been assimilated into Chasu nouns can be viewed in table 4.10 below. Loanwords take the same inflections and occupy the same syntactic slots as in the host language. It should be noted that some nouns are borrowed first from English to Swahili and thence from Swahili to Chasu. This especially applies to the category of cultural borrowings. In the 2nd and 3rd column of table 4.10, borrowed words from Swahili but not from English are given in bold.

Table 4. 10: Assimilated borrowed nouns in Chasu

N/C	Chasu	Swahili	English	Ultimate source
5/6	I-gita/Ma-	Gitaa/Ma-	guitar	English
5/6	I-gari/Ma-	Gari/Ma-	car	Hindi
5/6	I-kampuni/Ma-	Kampuni/Ma-	company	English
5/6	I-darasa/Ma-	Darasa/Ma-	class	Arabic
5/6	**I-kanisa/Ma-**	**Kanisa/Ma-**	church	Swahili
9/10	Timu/-	Timu/-	team	English
9/10	Shule/-	Shule/-	school	German/Hebrew?
9/10	Thieta/-	Thieta/-	theatre	English
9/10	Biashara/-	Biashara/-	business	Swahili/Arabic?
9/10	**Tangawizi/-**	**Tangawizi/-**	ginger	Swahili
9/10	Sekondari/-	Sekondari/-	secondary	English
1/2	Mw-alimu/Va-	Mw-alimu/Wa-	teacher	Arabic
1/2	**M-fanyakazi / Va-**	**M-fanyakazi /Wa-**	worker	Swahili/Arabic?
1/2	M-sabato/Va-	M-sabato/Wa-	sabbath keeper	Hebrew/English?

There are very few loanwords in Chasu borrowed direct from English. These are like, *m-masia* 'of messiah', *m-diakoni* 'a deacon', *presheni* 'operation', *digirii* 'degree', *i-diripu* 'drip feed'. Some of these words have equivalent terms in Swahili, which are *upasuaji* 'operation', and *shahada* 'degree' used in formal settings.

In the data, words from the Swahili noun classes 1/2, 5/6, and especially 9/10 are prone to borrowings compared to other noun classes. In contact between Swahili and English, noun class 9/10 of Swahili receives more nouns compared to the other noun class; likewise Chasu accepts nouns extensively in this class from Swahili and English. Those words, which are normally allocated to noun class 9/10 in Swahili borrowings from English, happen to occupy the same noun class when they are either borrowed into Chasu directly from English or from Swahili as the intermediate language. They are easily integrated into the Chasu noun prefix system, as some of them resemble each other either in singular or plural or both, as in noun class 9/10. It should be noted that words in noun class 9/10 has zero allomorphs for their nominal prefix. One can distinguish

whether they are singular or plural through their subject prefix, which is attached to the verb as an agreement. Words borrowed from English into Swahili are frequently slotted into noun class 9/10, and are easily borrowed into Chasu as they do not change their nominal prefix. However, it is only through their recurrence that one can tell if borrowed nouns in noun class 9/10 are loanwords and not single-word switches.

The data shows that verbs are also frequently borrowed compared to other grammatical categories. However, unlike nouns, verbs are not as easily integrated into the Chasu verb structure as nouns. Like other Bantu languages, Chasu and Swahili have a concordial system which involves agreement between a controller noun phrase and the verbs or adjectives. Both Swahili and Chasu have a subject prefix attached to a verb that has to agree with the noun class, which makes a noun phrase as in columns of table 4.8 of Swahili noun classes and table 4.9 of Chasu noun classes above. When a verb is borrowed from Swahili into Chasu, it acquires the subject prefix of the recipient language. Even though Chasu and Swahili are both agglutinative languages, their system of inflection for verbs is not only complex but rather different, as schematized below. Swahili has a pre-verb category that carries negations, agreement markers, tense and aspect marking, and object and relative markers (optional). The post-verb stem carries items like an imperative plural marked by the suffix **–ni**, a final vowel which carries the mood suffix **–a** (indicative) or **–e** (for imperative or subjunctive). Sometimes the verb stem may be followed by derivations which are either causative reciprocal, applicative, stative or passive form (Vitale 1981:13-14).

Table 4. 11: Swahili verbal structure

(Neg)	Subject affix	Tense	(relative affix)	(Object affix)	STEM	Post stem materials

Source: Vitale 1981:14

The pre-verb and post-verb stems of the Swahili verbal structure can be exemplified as in 2 and for Chasu verbal structure as in 3 below. Table 4.12 is adopted from Mreta (1998)

Example 2

(a) Ha-ta-m-pik-i-a
NEG-T/A FUT-OP3SG–VB-APPL-FIN
Not/(S)he-will-him-cook-for
'She will not cook for him'

(b) Wa-na-o-pig-an-a
SP3PL-T/A PRES-RLP-VB-RECP-FIN
They-are-who-fight-each other
'Those who are fighting each other'

(c) Ni-Ø- m-pig-e
SP1SG-T/A-OP3SG-VB-FIN/SUBJ
I-Ø- him/her- hit-should
'Should I hit her?'

(d) M-Ø-pig-e-ni
OP3SG-T/A-VB-IMP-PL
Him-Ø-hit-should-you
'You should hit him'

Table 4. 12: Chasu verbal structure

	PRE-VERBAL BASE DOMAIN									
Slot names	Pre-initial	Initial	Post initial	Formatives						object
Slot fillers	Neg 1 & Hortative	SMP	Neg2	T/A1	T/A2	T/A3	T/A4	T/A5	Neg	OPM
Morphemes occurring in each slot	Te- Na-	SPM1 SG/PL SPM2 SG/PL SPM1-16 Ku-INF	-esi-	(e) L H e- a-	-ka- -ki- -ki-	-ne- -na- -na-	-re- -re- -ra-	-cheri- -cheri- -ronga-	-sa-	OPM1SG/PL OPM2SG/PL OPM1-16 REM -ku-

VERBAL BASE & POST-VERBAL- BASE DOMAIN										
Root extensions									Final elements	Post-final
Root	Applicative	causative	intensive	Reciprocal	stative	inversive	passive	complex	FIN	PLVOC
	-i- -ir- -il-	-sh- -ish- -iž- -ež-	-ish-	-an-	-ik-	-uk-	-w- -iw-	- ižw- -any- -anyik-	-a -a -e -e -ie -ie	-eni

Key- for abbreviations

SPM1 SG/PL- subject pronoun marker 1st person singular/plural
SPM2 SG/PL- subject pronoun marker 2nd person singular/plural
SPM1-16- sprefix for noun classes 1-16
INF-Infinitive
OPM1 SG/PL- object pronoun marker 1st person singular/plural
OPM2SG/PL- object pronoun marker 2nd person singular/plural
FIN- final
PLVOC- pluralising final marker for object or subject prefix.
REM- relative marker
T/A- tense or aspect marking
OPM1-16 object prefix for noun classes 1-16

Example 3

(a) ni-Ø-m-big-i-a
SP1SG-T/A-OP3SG-VB-APPL-FIN
I-PRES-him/her-hit-for
'I hit for him/her'

(b) te-tu-Ø-va-kund-ish-a
NEG/PRO-SP1PL-T/A-OP3PL-VB-INT-FIN
Not-we-PRES-them-love-so much
'We do not love them so much'

(c) Na-tu-mu-im-iž-e-
Hortative-SP1PL-OP3SG-VB-CAUS-SUBJ
Let-we- him/her-cultivate-cause
'Let us make him cultivate'

Due to the complexity of the Chasu verbal structure, as we can see in table 4.12, when a verb is borrowed from Swahili or English, it normally receives the subject prefix inflection, which is common to Swahili and Chasu, or the infinitive marker *ku-*. Sometimes it receives the tense or aspect marker or other inflections, especially those with the same position in the word formation. Occasionally it is difficult to choose an inflection, since Chasu has many inflections compared to Swahili. The derivations occur in both Chasu and Swahili verb forms. Verbs of Arabic origin in Swahili rarely have derivational forms, because they do not concur with verbs of Bantu origin which have **–a** as the final vowel. Thus, except for such borrowed verbs, all borrowed forms take both prefix and the post-stem derivations. We can see in the example below how some of the borrowed words are inflected with Chasu pre-verb and post-verb stem inflections.

Example 4

(a) ve-ki-*chez*-i-a
SP3PL-T/A PROG-VB-APPL-FIN
They-were - play-for
'They were playing for'

(b) ni-Ø-*vami*-w-e
SP1SG-T/A-VB-PSV-FIN
I-PST-attack
'I was attacked'

(c) ni-Ø-m-*shukuru*
SP1SG-T/A-OP3SG-VB
I-PRES/CONT-him/her-thank
'I thank him/her'

(d) ve-ki-ra-*chez*-a
SP3PL-T/A2-VB-FIN
They-were-still-play
'They were still playing'

(e) e-Ø-ku-*ongoz*- a
SP3SG-T/A-OP2SG-VB-FIN
She/he-PRES/CONT-you (sing)-lead
'She leads you'

Example 4 (a) and (b) have pre-verbal aspects of subject prefix, tense or aspect as well as verbal extensions for applicative form and passive form respectively. Examples (c) to (e) have only the prefixes while (c) is one of the verbs which are borrowed from Arabic. Example (d) has a tense form *–ra-* 'still', found in Chasu verb formation but not in Swahili.

Though there is a strong tendency in Chasu to borrow lexical items from Swahili and English, it has been at times hard to ascertain whether certain nouns are borrowed or code-switched. Myers-Scotton (1993b:167) claims that scholars working on different data sets are concluding that 'morphological and syntactic linguistic subsystems are much more open to borrowing than previously acknowledged, and that not all lexical borrowing forms are equally integrated into the recipient language'. This statement applies to the study of Chasu, where core borrowed words which are allocated in noun class 9/10 have a zero allomorph for the prefix. As mentioned before, these noun classes are the same in Chasu and Swahili. When a borrowed word is slotted into this noun class in Chasu, it is difficult to assert whether the word is borrowed or switched, since no assimilation takes place.

***Numbers*:** In this study, numerical expressions are more frequently borrowed than any other word types. Table 4.13 below illustrates the numerals which were borrowed from Swahili and English and their frequency. The total frequency of borrowed numerals is more than half of the total numerical expressions used in the entire conversations. It needs to be emphasized that earlier Chasu had equivalents for all these terms, as tables 4.14 and 4.15 show below.

Table 4. 13: The frequency of borrowed numbers from Swahili and English

Borrowed numbers	English	Frequency
elfu	thousand	83
sabini	seventy	54
saba	seven	48
sitini	sixty	45
themanini	eighty	44
sita	six	42
hamsini	fifty	40
mia	hundred	39
tisini	ninety	38
moja	one	35
tisa	nine	33
-a kwanza	*first*	*28*
-a saba	*seventh*	*19*
arobaini	forty	17
four		**9**
ishirini	twenty	9
nne	four	8
mbili	two	9
one		**6**
-a sita	*sixth*	*5*
thelathini	thirty	4
tano	five	4
tatu	three	4
three		**3**
six		**2**
two		**2**
sifuri	zero	1
Total		**631**

Out of a total 1250 numerical expressions in the corpus, 619 were from Chasu, while 609 were borrowed from Swahili and 22 were from English (bold in table 4.13). Borrowed numerical forms comprise 579 cardinals and only 52 ordinals (italicized). From Swahili, ordinal borrowed numbers included *–a kwanza* 'first', *-a sita* 'sixth' and *-a saba* 'seventh' from single digits. Chasu speakers borrow cardinal numbers *sita* 'six', *saba* 'seven', and *tisa* 'nine'; *ishirini* 'twenty' up to *mia* 'one hundred' and a*lfu* 'thousand'. These are the numerical terms which were also borrowed from Arabic via

Swahili. Speakers still use other Chasu single digits in their conversations, while tens (from twenty), hundreds, and thousands are referred to in Swahili. In this study, the Chasu word *igana* 'hundred' was only once used, and *ikumi* 'ten' was used twice. The remaining numerals were from Swahili, except the single digits *mwe,* 'one', *mbiri* 'two', *ntatu* 'three', *ne* 'four', *sano* 'five' and *mnane* 'eight'.

Borrowed numbers from Swahili are not assimilated to Chasu structure phonologically, syntactically or morphologically, even though they occur frequently and seem to be socially acceptable judging from their frequency. There are speakers, especially younger ones, who hardly remember Chasu numbers. Though borrowed numerical forms are not absolutely integrated, they have been grouped with other core borrowing forms because Chasu has its own numerical expressions, as shown in tables 4.14 and 4.15 below, and the borrowed ones are considered to be used for reasons of prestige.

Table 4. 14: Numerical system: single digits in Chasu

Chasu	Swahili	English
mwenzu	moja	one
mbiri	mbili	two
ntatu	tatu	three
ne	nne	four
sano	tano	five
mtandatu	sita	six
mfungate	saba	seven
mnane	nane	eight
kenda	tisa	nine
ikumi	kumi	ten

Table 4. 15: Tens and hundreds in Chasu

Chasu	Swahili	English
makumi meri	ishirini	twenty
makumi matatu	thelathini	thirty
makumi mane	arobaini	forty
makumi masano	hamsini	fifty
makumi mtandatu	sitini	sixty
makumi mfungate	sabini	seventy
makumi mnane	themanini	eighty
makumi kenda	tisini	ninety
igana	mia	hundred
kiku	alfu	one thousand
laki	laki	one hundred thousand
milioni	milioni	million

In Chasu, cardinal numbers occur as nouns (n=580) as in sentence 5(a) and quantifiers (n=40) as in 5(b) in examples below.

Example 5

(a) Nimogiwe mwaka *alfu moja mia tisa na arobaini na saba*
'I was born in 1947'

(b) Nina vana *saba*
'I have seven children'

When they function as quantifiers, the numbers *mwenzu, mbiri, ntatu, ne* and *sano* show the morphological agreement with the noun class prefix of the noun phrase ,except for noun class 9/10. The numbers *mtandatu, mfungate, mnane, kenda* and *ikumi* do not show morphological agreement despite the change of the noun class or the plurality of the nouns. We can view this in table 4.16 below:

Table 4. 16: Noun class and Chasu numerical system

Noun class/ numbers	Numeral one	Numeral two to five	Numeral six to ten	
1/2 M-/Va-	M-ntu m-mwenzu 'one person'	Va-ntu ve-ri (2) Va-ntu va-tatu (3) Va-ntu va-ne (4) Va-ntu va-sano (5)	Va-ntu mtandatu (6) Va-ntu mfungate (7) Va-ntu mnane (8) Va-ntu kenda (9) Va-ntu ikumi (10)	Person(s)
3/4 M-/Mi-	M-ti m-mwenzu 'one tree'	Mi-ti mi-ri (2) Mi-ti mi-tatu (3) Mi-ti mi-ne (4) Mi-ti mi-sano (5)	Mi-ti mtandatu (6) Mi-ti mfungate (7) Mi-ti mnane (8) Mi-ti kenda (9) Mi-ti ikumi (10)	Tree(s)
5/6 I-/Ma-	I-jembe i-mwe 'one hoe'	Ma-jembe me-ri (2) Ma-jembe ma-tatu (3) Ma-jembema-ne (4) Ma-jembe ma-sano (5)	Ma-jembe mtandatu (6) Ma-jembe mfungate (7) Ma-jembe mnane (8) Ma-jembe kenda (9) Ma-jembe ikumi (10)	Hoe(s)
7/8 Ki-/Vi-	Ki-ogwe ki-mwe 'one potato'	V-i-ogwe vi-ri (2) Viogwe vi-tatu (3) Vi-ogwe vi-ne (4) Vi-ogwe vi-sano (5)	Vi-ogwe mtandatu (6) Vi-ogwe mfungate (7) Vi-ogwe mnane (8) Vi-ogwe kenda (9) Vi-ogwe ikumi (10)	Sweet potato(es)
9/10 N/N	Shuke mwenzu 'one cloth'	Shuke mbiri (2) Shuke ntatu (3) Shuke ne (4) Shuke sano (5)	Shuke mtandatu (6) Shuke mfungate (7) Shuke mnane (8) Shuke kenda (9) Shuke ikumi (10)	Cloth(es)

Table 4.16 illustrates that when nouns take their plural forms, as in column 3, traditional Chasu numbers *mbiri,* 'two', *ntatu* 'three', *ne* 'four' and *sano* 'five' take the morphological agreement from the noun class, except for noun class 9/10. Column 4 illustrates that numbers *mtandatu*, 'six', *mfungate*, 'seven', *mnane* 'eight', *kenda* 'nine' and *ikumi* 'ten' do not change despite their plurality. When the numbers in column 4 are

replaced by borrowed Swahili/Arabic counterparts *sita* 'six', *saba* 'seven' and *tisa* 'nine', they likewise have a zero prefix form, since this is what occurs in all three languages concerned, Arabic and Swahili as well as traditional Chasu. Likewise numbers *ishirini* 'twenty' up to *mia* 'hundred' and *alfu* 'thousand' do not take any agreement, as they behave like other nouns which are categorised into the noun class 9/10, which has a zero allomorph as noun class prefix.

Numbers which are borrowed from English are considered in this study as non-established forms, because they do not occur frequently. They can be considered singly as lexemes in code-switching, since they do not recur, are not assimilated, and are used by few educated trilinguals. Of all 22 English borrowed numerical expressions, 21 occurred amongst highly educated speakers, while one token was from a less educated speaker. Though they are few compared to Swahili borrowed numerical expressions, their number is likely to increase as English and Chasu come into contact more directly. All numerical forms from English were borrowed when speakers were expressing their level of education, particularly in secondary school; see examples 6 (a) - (d) below. Thus, due to their functions in the sentence structure, these numbers happened to occur more as part of a compound noun than as numerals on their own.

Example 6

(a) Kipindi nekishoma **form three** mpaka **form four** hakatea mgomo
'When I was doing form three to form four there was a strike.'

(b) …nikatonga Muheza sekondari **form one** na **two, form three** na **four** Parane
'... I went to Muheza secondary for form one and two, three and four in Parane'.

(c) … nikaronga mtihani wa **form six** nikafaulu
'I did and passed my form six examination'.

(d) Vana vangu vose veshomie mpaka **form four**
'All my children were schooled up to form four'.

Apart from such numerical forms borrowed from English, numbers occurred when the respondents discusses or mentioned the following topics: dates of birth and ages, marriage ceremonies, historical events such as festivals or catastrophic phenomena like the coming of locusts, dates of schooling, other levels of education attained, number of children and other possessions like animals (goats and cows) etc.

Generally, the number of borrowed numerals is surprisingly large, with above half of all used numerical terms in the corpus. Though borrowing instances are quite high, these forms are not found in any Chasu written texts. In these texts, people could easily access traditional Chasu numerals. In speech, however, speakers particularly youngsters do not use them much. This situation suggests that, with time, more changes will occur in the use of borrowed numerals, especially when even more people are in contact with Swahili through education, travelling and business.

The study of numerals in Chasu raises several critical issues. Numbers in this study are borrowed from Swahili words, themselves being borrowed from Arabic. They have been categorised into the group of core lexical borrowing because numerals are not a new phenomenon in this society. Like single-word switching, neither cardinal nor ordinal borrowed numerical expressions are assimilated into Chasu structure; hence they behave like single-word switches. Lack of assimilation goes contrary to the argument of Sankoff, Poplack and Vanniarajan (1990:74) that "the clearest cases of B (borrowing) forms are those forms showing phonological, morphological and syntactic integration into the ML (Matrix Language), but phonological integration may not always be complete". This may occur in other studies, but we cannot observe the same kind of structural integration in Swahili and Chasu. Nevertheless, borrowed numbers are predictable as to their recurrence and they seem to be accepted by the speakers, since the data shows that they are the most borrowed items, especially the 'tens' among young age speakers. Owing to these traits it is difficult to ascertain whether borrowed numerical expressions are loan words or single-word switches. This makes the recurrence and its acceptability among the language users more important criteria than assimilation in distinguishing lexical borrowing from code-switching.

4.2.2 Code-switching and constraints

In the next two sections I characterise code-switching from a structural perspective, as a background to an analysis of sociolinguistic variation in code-switching in section 4.3.

While in some communities code-switching is considered an exceptional case, in most bilingual and multilingual societies it is becoming not only a norm but a very salient socio-cultural phenomenon. As is now well recognised, since the early research of Poplack (1980), code-switching is not just a "performance error" (Myers-Scotton 1993b:50), caused by a lack of ability to sustain an ongoing conversation. Rather, people who engage in code-switching are bilinguals who are fluent and proficient in both of their languages. Hence code-switching is not usually motivated by a struggle to select words from any of the participating languages. Moreover, the utterances in code-switching, accord with the conventions of the grammar of the participating languages. During field work, though the conversations were initiated in Chasu and no questions were asked in Swahili, the respondents frequently switched to Swahili and English. Hence the corpus includes linguistic alternation within the sentence and beyond.

Previously, code-switching, especially intra-sentential switching was considered syntactically random and deviant rather than rule-governed (Weinreich 1953/1968). In his study investigating the syntax of Spanish/English code-switching, Lance (1975:143), as quoted by Redouane (2005:1922), concludes that 'there are perhaps no syntactic restrictions on where the switching can occur'. However, since then researchers have proposed various linguistic factors that operate to constrain code-switching and the syntactic constraints that restrict the environments where language alternation can occur. One such early, well-known suggestion was that in intra-sentential switching, code changes tend to occur where the syntax of two languages align. This principle is known as the Equivalence Constraint, after Poplack (1980:586), who proposed that:

> Code switches will tend to occur at the point in discourse where the juxtaposition of L1 and L2 elements does not violate a syntactic rule of either language, i.e. at points around which the surface structures of the two languages map onto each other. According to this simple constraint, a switch is inhibited from occurring within a constituent generated by a rule from one language which is not shared by the other.

According to Muysken (1987:123), in this principle switching from one language to another in the middle of a sentence is only possible if the linear order of sentences in both languages is preserved. In the Equivalence Constraints model, word order requirements of both languages are met at the S-structure. The model was postulated simultaneously with the Free Morpheme constraint (Sankoff and Poplack 1981:5) which predicts that 'a switch may not occur between a bound morpheme and a lexical item unless the lexical item has been phonologically integrated into the language of the bound morpheme'. These models were based on the Spanish/English bilingual speech of the United States.

Di Sciullo et al (1986) proposed a model of Government constraints in code-switching, which is based on X-bar theory, where the notion of syntactic relations relies on the 'head of a category'. The Government principle is based on the assumption that there is syntagmatic coherence in the sentence, and the head node governs the immediate constituent; i.e. 'X dominates Y if (if and only if) the first node dominating X, also dominates Y, where X is a major category (noun, adjective, verb, preposition)' (Romaine 1995:130). Di Sciullo et al claim that switching is only possible between elements that are not related by government, meaning that there will be no switches between elements that are lexically dependent on each other.

Possibly the most detailed constraint model of code-switching is the Matrix Language Frame (MLF hereafter) model of Myers-Scotton (1993a&b, 2002). It involves a contrast between content morphemes and functional (or system) morphemes. Central to the MLF model is the premise that, although both languages may participate in the

production of CS, they are unlikely to play equal roles. MLF model analysis involves two or more participating languages, which are the Matrix Language (hereafter ML) and the Embedded Language (EL). While the EL plays a lesser role of giving content morphemes in the code-switching utterances, ML is the core language which sets the morpho-syntactic frame in the mixed constituents and determines the morpheme order (Myers-Scotton 1993a & 1997:83). The mixed constituents contain content morphemes from both ML and EL, but the grammatical frame is from ML. The grammatical frame is defined by the morpheme order and system morphemes. System morphemes include all syntactically active inflections and most function words in bilingual constituents, which are normally from the ML and are defined as features [–thematic role receiver/assigner] and often [+quantification]. Myers-Scotton (1993b:6-8) adds that, in contrast with system morphemes, content morphemes are divided according to their lexical category, which either receives, thematic roles (most nouns and adjectives) or assigns them (most verbs, some prepositions). Though speakers have to be bilingual to perform code-switching, competence in the matrix language is particularly crucial, since it supplies the morpho-syntactic frame of the mixed constituents. Subsequently, the node 'sentence' in the model has been replaced by bilingual 'CP' (projection of complementizer), which is considered to be the best unit of analysis for examination of any contact phenomena (Myers-Scotton 2002:54). It only becomes a bilingual constituent when it contains three code-switching constituents, namely: (a) Mixed or ML + EL with morphemes from two or more languages, (b) ML islands, constituents with morphemes solely from the ML and well-formed according to the ML grammar, and (c) EL islands, constituents with morphemes solely from EL and well-formed according to the EL grammar.

In my database, it is evident that Chasu is the matrix language while Swahili and English are the participating embedded languages. It is important to note the difference between this study and that of Myers-Scotton. In the Kenyan studies (Nairobi) reported by Myers-Scotton (1993a), Swahili is a matrix language, while in Zimbabwe, Shona is the matrix language but like Swahili and unlike Chasu it is itself a lingua franca. Examples showing how this model works in my Chasu database are discussed, along with intra-sentential code-switching, in section 4.2.3(c) below.

4.2.3 Code-switching in Chasu

Generally the forms of code-switching in the Chasu data-base constitute a hierarchy ranging from the whole sentences, phrases and clauses, and other chunks of discourse of single-words which can be inserted into the grammatical structure of a ML. At an initial level, these can be characterised in terms of tag-switching, intra-sentential and inter-sentential switching. These terms are defined in their respective sub-sections, and their relative frequency is given in table 4.17.

Table 4. 17: Frequency of different types of code-switching

Types	**Tag switching**	**Inter-sentential switching**	**Intra-sentential switching**	
			Single words	Intra-clausal switches
	3	59	72	220
Total	**354**			
%	0.84	16.66	20.33	62.14

(a) Tag switching

Tag switching involves the insertion of a tag from one language into an utterance in another language. The term 'tag' covers elements like discourse markers, and affirmative and negative particles. Romaine (1995:122) comments that, since tags are subject to minimal syntactic restrictions, they may be inserted in a number of points in a monolingual utterance without violating syntactic rules. In this study there were very few tags. Table 4.17 shows only 3 tokens from English, which constitute 0.84% of all the switched utterances.

Example 7

(a) Q: Udima kunielezia maisha ako ose hangi umogiwe mwaka ani?
A: **Okay!** mi nimogiwe mwaka elfu moja mia tisa hamsini na tano

Q: 'Can you tell me about your life, when were you born?
A: Okay! I was born in one thousand nine hundred and fifty five'

(b) Q: Elezia familia yako na ndima žako.
A: **Okay,** nina… kwa kimpare?

Q: 'Explain about your family
A: Okay, I have… in ki-Pare/Chasu?'

(c) (Laughing) **Anyway**, nikasova nafasi ya kuendelea na shule
'Anyway, I had no chance to continue with school'

In this study, all three tokens of the tag switching were inserted at the beginning of the sentence. In examples 7 (a) and (b) the word **okay** is inserted in the sentence for affirmative function in the discourse, while the word **anyway** in example (c) leads into diverging from the previous idea within the conversation.

(b) Inter-sentential switching

This type involves a language alternation between sentences. It involves producing one or more sentences in one language before producing a sentence from the other language(s) in use within a speech event (Myers-Scotton 1993a:3-4). They may occur within and across speakers' turns, and they are thought to require greater fluency in both languages. This is because major portions of the utterance must conform to the rules of both languages. As we can observe in table 4.17, there are 59 portions of utterances in inter-sentential switching which make 16.6% of all the switched utterances. This type of switching involves monolingual sentences from either Chasu or Swahili, and sometimes some embedded phrases from English as in the example below:

Example 8

Niendelea na shughuli žangu ža biashara hamwe na kimo ni kazi mbaha Neikundie zaidi mira shughuli neikundie sana ni biashara na kimo nicho vintu nekikundie sana. *Mambo mengine ya dini nayo najishughulisha nayo, mimi ni mkristo muadiventista msabato na kanisa langu ni* **Manka SDA church** *liko katika mtaa wa Myamba* **North East Tanzania Conference**

'I continue with business activities and farming is the main work I like but both business and farming are the works I like most. I do other things related to religion as well, I am a Seventh Day Adventist and my church is Manka SDA church which is in Myamba ward in North East Tanzania Conference'

In example 8, there are two sentences within a conversation from either language. Initially the speaker commences with a Chasu sentence, and then switches to a Swahili sentence which has accommodated two phrases from English. The first monolingual sentence is structurally independent, as Chasu grammatical rules are observed. In the second sentence Swahili is the ML with two embedded noun phrases from English.

In this type of switching it is sometimes difficult to establish a demarcation between Swahili and Chasu as to which language is ML or EL, as a respondent may directly respond in Swahiliwithout starting with a Chasu utterance, as we can see in example 9 below:

Example 9

Q: Ni kintu ani wekikundie he maisha ako kukironga zaidi ya biashara žako yani hangi chekuburudisha wekizihirwa?

A: *Ninapenda sana kujisomea, ninatamani sana kujiendeleza kusoma japo umri wangu sasa umesonga lakini ni kintu ambacho kinaleta manufaa ya mtu.*

Q: 'What thing do you like most to do more than your business, something which can entertain you?'

A: 'I like reading so much, I highly wish to improve my education level, though my age has gone far but it is the thing which benefits an individual'.

Though the question is asked in Chasu in example 9 above, the respondent replied in Swahili, which is somewhat unusual. Although Chasu is the ML in much of the interviews, there are chunks of utterances in which Swahili is the matrix and Chasu phrases are inserted into it, as in the example below:

Example 10

Kwa sababu sisi ni wafanyabiashara **tuzana na majambazi** *ni tukio hivi karibuni majambazi wamevamia gari ambalo tunasafiri nayo lakini nusura mimi nashukuru Mungu tu siku hiyo sikuwapo hapo lakini wenzangu walinyang'anywa pesa zote …*

'Because we are businessmen, **we meet robbers**, there is a recent incident when robbers hijacked the bus which we always travel on but luckily I thank God, that day I was not there but my fellows were robbed of all their money….'

Utterances such as that in example 10 indicate that, to some Chasu speakers, the contact between Chasu and Swahili is characterised by stable bilingualism and alternation between each possible ML. Respondents might use such stable patterns to convey social meaning. That is, such bilinguals may want to identify themselves with Swahili speakers and urban ways of life through showing their competence in Swahili. They may want to project a sense of shared, dual identity and status with the interlocutor, especially if the latter is also a bilingual and/ or holds higher socio-economic status than the speakers. However, there are limits to switching for reasons of prestige, as can be seen by examining switches from English. There were no inter-sentential switches involving English despite its high status. This may be due either to its distancing functions or to the speaker's lack of adequate proficiency in the language.

(c) Intra-sentential switching

This involves switching forms 'within the clause or sentence boundary' (Romaine 1995:123) and sometime includes mixing within word boundaries, where there is an insertion of inflectional morphology from another language. This type requires the most fluent bilinguals, who are able to control two or more linguistic systems simultaneously**.** In this context, I have categorised two main levels of intra-sentential switching based on Myers-Scotton's (2006:253-266) conceptualization of singly occurring words from phrasal level code-switches**.** Thomason (2001:136) comments that single words and short phrases are the most common code-switched elements. In this study, phrases are the element most frequently switched, up to 62.14%, according to table 4.17. The MLF model best applies to intra-sentential switching, as we will discuss below.

Single words: In code-switching, the occurrence of single lexemes is different from that in lexical borrowing. Unlike from established lexical borrowings, single word switches involve the insertion of words which have not been fully assimilated into the structure of the ML, and do not necessarily recur. In this study, there are 72 (20.33%) single-lexeme switches of different word categories, 19 from English and 53 from Swahili. Based on Myers-Scotton's (2006:254-255) categorisation of single word switches, single lexeme switches in this sudy involve two major groups; those which are partially assimilated (as in example 11 below) and those which are not assimilated (as in example 12).

The partially assimilated group often involves verbs which have received the system morphemes (inflectional affixes) from the ML, while the content morphemes (word roots) have remained in the EL form. Examples of these are in 11 below.

Example 11

(a) Shule **že-kiho-scattered** sana.
Schools SP10-T/A-scatter much
'Schools were so very scattered'.

(b) Nikatonga Singida **ku-upgrade** mwaka elfu mbiri na mwe.
I- went Singida INF-upgrade year thousand two and one.
'I went to Singida for upgrading in year two thousand and one'.

(c) kale twekitonga *ku-ogelea* uko
Long time ago we were going INF-swim there
'Long time ago we were going to swim there'.

(d) Vangi ni vaalimu Mfumwa e-*Ø-va-bariki*
Some are teachers Lord SP3SG-T/A/PRES-OP3PL-bless
na vangi *ve-Ø-ra–som-a*
and some SP3PL-T/A1-TA2-study-FIN
'Some are teachers, the Lord is blessing them, and some are still studying'.

(e) ... shughuli žekunka ka-**income** keridhisha
Activities give DIM-income that satisfies
'...Activities which give you a satisfying income'.

For instance words like ***ze-ki-ho*-scattered** in 11 (a) and ***ku*-upgrade** in (b) involve switches from English. ***Ku-*** *ogelea* in (c), ***e-va****-bariki* and ***ve-ra****-soma* in (d) are switches from Swahili. Both have retained their verb stems from the ELs, but the bound morphemes are from the ML, except 11(a), in which the English past tense was treated as part of the root in this context. In this study, there are very few nouns from English which show assimilation to Chasu, as in example 11(e). These examples have the characteristics of nonce borrowing, in the sense that they involve lone lexical items, especially with content words. Though they are not fully assimilated, like established loan words, they assume some morphological and syntactic aspects of the recipient language. At the same time they are neither recurrent nor widespread.

Another group of single-word switching involves isolated structures of EL forms, in the sense that they have no attached elements from ML. These involve examples as in 12 below:

Example 12

(a) ….nikatonga kujifunza **nursing** muda wa mieži mitatu nikashiga
'….. I went for nursing skills for three months'.

(b) Hakatea **secretary** akaza akamwira iti mi neki mwanafunzi.
'There was a secretary who came and told her that I was a student'.

(c) Tukashoma tukapata **credit** žose.
'We read and got all credits'.

(d) **Graduation** yangu neki yedi ambu twekina na **pastor** mwe.
'My graduation was nice because we had one pastor'.

(e) Kurerehia *burudani* sa ižo.
'Watching entertainments like those'.

(f) He kikundi cha vache hena *masomo* a *ujasiriamali*
'In the women group there are entrepreneurship lessons'.

In these examples, there are nouns like **nursing** in (a), and **secretary** in (b), **credit** in (c), **graduation** and **pastor** in (d) from English. Swahili nouns include *burudani* 'entertainment' fin (e), *masomo* 'lessons', and *ujasiriamali* 'entrepreneurship' from (f). These examples are fully EL forms; no assimilation has taken place at all. Unlike the examples in 11 above, the examples in 12 are 'bare forms', in the sense that 'they do not receive any inflections that would make them well-formed in the language that supplies the morphosyntactic frame (ML)' (Myers-Scotton 2006:255). Like other forms of code-switching, their recurrence is not predictable.

According to English word order, the adjective precedes the noun, while in Chasu it is the noun which comes first. Although switched nouns from English have not been assimilated, they have been slotted into the head-first word order of the Chasu noun phrase (noun first, then modifier or quantifier), as well as the noun class. Although the nouns themselves have no Chasu prefix, their agreement markers do. For instance, **graduation ya-ngu** 'my graduation' in (d) and **credit ž-ose '**all credits' in (c) are located in noun class 9/10, while **secretary a-kaza '**came a secretary' in (b) and **pastor m-mwe** 'one pastor' in (d) are in noun class 1/2. The nouns are all followed by a modifier, quantifier or verb in agreement with the noun class from ML. These forms substantiate the essential difference between ML and EL in my data set. Furthermore, Myers-Scotton (1993b:206) has suggested that lexical borrowings and single word switching have a close relationship, as both forms are subject to the morphosyntactic procedures of the recipient language. In their initial stage, they are all code-switched materials. This, too, is supported by examples in this section.

***Phrasal switchings*:** These involve switching different levels of constituents, ranging from phrases to large chunks of clauses formed within a sentence. In my data, there are a total number of 220 (62.14%) phrasal level switchings, of which 187 are elements from Swahili and 33 constituents are from English. In this study, these are the most switched elements compared to single-word and sentential structures. We can view some of the occurrences below.

Example 13

(a) Mpaka iki niendelea nekioka iti [$_{COP}$ ni] [$_{NP}$ *muuguzi* **grade B**]
'Up until now I continued working as a nurse in grade B'

(b) [$_{NP}$ Kimo] [$_{COP}$ ni] **[**$_{NP}$ *uti wa mgongo wa kitaifa*]
'Agriculture is the backbone of the nation'

(c) Hena mwe eronga ndima [ya] [$_{NP}$ **hotel management and Tourism]**, ungi eguhiwe
'There is one who works with Hotel management and Tourism, one is married'

(d) [$_{MC}$ *Niliingia thieta saa kumi na mbili*,], [$_{MC}$ nifumie hala thieta sa sita ža kio]
'I entered the theatre room at six o'clock, I came from the theatre at twelve midnight'

(e) [$_{NP}$ Mi] [$_{VP}$ nimogiwe] [$_{AdP}$ *mwaka wa alfu moja mia tisa arobaini na tisa*.]
'I was born in 1949.'

Example 13 (a) *muuguzi* **grade B,** and (b) *uti wa mgongo wa kitaifa* are noun phrases which occur after the copular *ni* functioning as subject complements. The switched element from English in example (c) is a noun phrase made up of 2 conjoined nouns; it functions as object complement. In example (e) a switched element is an adverbial phrase which modifies the verb to indicate year of birth. Switches in examples (a), (b), (c) and (e) follow the word order of Chasu**.** Example (d) is a constituent of two independent clauses, starting with a switched independent clause from Swahili followed by another independent clause from Chasu; they are connected by a comma/pause. Both clauses are composed with verb phrases as the head category. This structure is what Gardener-Chloros (1991:175) classifies as a disjointed switch, where a constituent is made up of two different language segments, represented in the form of independent clauses.

Thus, from this study the most likely syntactic constraints for code-switching are as follows:

(a) As a subject complement after the linking verb or copula *ni* as in examples 13 (a) and (b) above.
(b) As adverbial construction after a verb phrase as in example 13 (e);
(c) As a noun phrase after the prepositional phrase to express location as in:

14. [$_{VP}$ Nikaronga] [$_{NP}$ ndima] [$_{PP}$ he] [$_{NP}$ *idara ya Malaria kwa muda*]
'I worked in the department of Malaria for a while'.

15. [$_{VP}$ Eronga][$_{NP}$ ndima] [$_{PP}$ he][$_{NP}$ *kitengo cha Sayansi na Elimu ya Juu.*]
'He works in the Science and Higher Education section'.

(d) As the main clause after the subordinate clause as in:

16. [$_{SC}$ Baada ya kuvona iti ndima zikaa muda mueza kidogo], [$_{MC}$ *nikaona nifanye kazi ile ambayo itamzidi mwenzake.*]
'After realising that two jobs could take time, I decided to do the job which surpassed the other'.

17. [$_{SC}$ Baada ya kujenga kwa muda wa miaka mitatu], [$_{MC}$ *nikaona kwamba ile kazi haitanifaa.*]
'After building for three years, I realised that the job did not benefit me'.

(e) As the noun phrase after the main verb as in:

18. [$_{VP}$ Akakenja] [$_{NP}$ *shule chuo cha usafirishaji*]
'He finished schooling in the college of transport'.

19. [$_{NP}$ Vana vangu niho] [$_{VP}$ vakenja] [$_{NP}$ *shule form four*]
'My children have just finished *school in form four*.

(f) As a noun phrase at the beginning of a sentence

20. **[**$_{NP}$ **Graduation** yangu] [$_{VP}$ neki yedi]
'My graduation was good.

As mentioned before, Swahili and Chasu have SVO word order, so it is easier to insert words, phrases or clauses from one language to another, since the word order is the same. Insertion of phrases from Swahili involves changing the language of lexis, but not word order position. Phrases from English follow the word order of the ML, as in example 20, i.e. the possessive pronoun (from Chasu) follows the English noun.

Generally, the study has shown how the MLF model can be applied in Chasu/Swahili or Chasu/English speech, especially at the level of words, phrases and clauses. However, at the sentential level it has been difficult to apply the MLF model. Code-switching at the sentential level involves alternation of sentences which are grammatically independent from each other, and where the ML/EL roles seem to change regularly.

4.3 Code-switching and lexical borrowing in relation to social factors

Rampton (1995, 2007) and Pujolar (2004) describe the effects which the use of multiple language varieties has on class, ethnicity, gender or other identity positions. Whether a person has the ability to participate in a conversation involving code-switching depends on that person's linguistic repertoire. This repertoire may correlate positively with certain demographic features. For example, in some societies, without a certain level of education it is unlikely that a person will be able to speak the linguistic variety associated with political and social economic power in the community.

One micro level aspect of code-switching, which associates the degree of code-switching usage with demographic variables, has been studied by Poplack and her associates (Poplack et al. 1988). They analysed speakers with different socio-political profiles in French Canadian communities, to ascertain who were the most frequent users of either borrowed English lexemes or code-switching. In addition to such factors as social class, they considered proficiency and community attitudes. They concluded that "the norms of the community override individual abilities" (Poplack et al. 1988:97-8) as the best predictor of an individual's use of code-switching. What one's peers in the community do is more influential than demographic variables or even one's own linguistic proficiency. The degree of contact, however, overrules the other factors. In elaborating this, Thomason (2001:66) argues that the higher the intensity of contact, the higher the possibility of various kinds of interference among bilinguals. She includes three main social aspects which can help to extrapolate the concept of intensity of contact:

1. The duration of contact period of the two languages in contact, as the longer the time, the higher the interference of the structure of the languages.
2. Size of communities: if one of two groups in contact is larger than the other, all other things being equal, the smaller group's language is more likely to acquire features from the larger group's language than if the two groups are roughly equal in size. A mitigating factor, however, is the relative prestige of a group.
3. The more socioeconomic dominance one of the groups exerts, and the higher its prestige, the more likely it is that members of the subordinate group will adopt features from the dominant group's language.

However, contact linguistics has traditionally paid less attention to internal diversity of speakers in contact situations, treating the groups in contact as relatively homogeneous. Since this thesis is ultimately of a sociolinguistic nature, it will investigate whether there are differential sub-groups of speakers in the contact situation. The following section examines how age, gender, social class and education level are directly related to the scope of borrowing and switching codes in the trilingual Chasu community.

4.3.1 Age groups

In this study there are three main age groups as mentioned earlier in chapter 3. These groups are composed of 23 young speakers from 21-40, 21 speakers from the middle-aged group ranging from 41-60 and 13 old-aged speakers of 61 years and above. It was mentioned earlier that the population in this society is mostly composed of young people, followed by the middle-aged and then old-aged, who are relatively few. This is portrayed in the number of speakers per age group. The occurrence of code-switching and lexical borrowing in relation to age difference is schematized in table 4.18

Table 4. 18: Age groups in relation to lexical borrowing

	Age 21-40 n=23 (40.3%)		**Age 41-60 n=21 (36%)**		**Age 61+ n=13 (22.8%)**		**Total**
Lexical borrowings	**core**	**cultural**	**core**	**cultural**	**core**	**cultural**	
Verbs	67	18	43	35	30	4	
Nouns	97	261	100	206	39	117	
Adverbs	4	-	16	-	-	-	
Adjectives	-	-	5	-	1	-	
Conjunctions	-	-	6	-	-	-	
Preposition	-	-	-	-	-	-	
Pronouns	-	-	-	-	-	-	
Numerals	269		240		122	-	
Total	**437**	**279**	**410**	**241**	**192**	**121**	**1680**
Grand total	**716**		**651**		**313**		**1680**

Table 4.18 indicates that in this study there is a total number of 1680 borrowed words, comprising cultural and core borrowings. As mentioned earlier in section 4.2.1 (cf. tables 4.4 and 4.5), all tokens were counted, including repetitions of the same item. In relation to age difference, the data in table 4.18, when added up, shows that young speakers use 716 borrowed words, followed by 651 by the middle-aged speakers, while the old-aged informants use 313 of all the borrowed forms. In each group the core borrowings outnumber cultural borrowings. This appears to be a slightly unusual state of affairs, since we would expect the main reason for borrowing to be related to new cultural phenomena or the prestige associated with cultural forms of higher status external groups in a contact situation. The explanation lies in the grouping of the numerals with core

borrowings, since they had exact equivalents in earlier forms of Chasu (as discussed earlier). If we factored out the numerals (which are frequently occurring items) then the results for core:cultural borrowings are reversed as follows: Young 168:279; Middle 170:241; Old 70:121. As there are fewer old speakers than young and middle-aged speakers it cannot be directly concluded from table 4.18 that older speakers use fewer loans. The totals in table 4.18 are raw figures which need to be tempered by (a) the length of each interview against which each subject's quantity of borrowing can be measured, and (b) the number of speakers per group. It was therefore decided that pairwise t-tests be conducted in EXCEL between the age groups, taking these factors into account. To enable comparisons between individuals, the number of borrowings was divided by the total length of time of the interview in minutes and then multiplied by 5 – i.e. a standardised measure of borrowings per 5 minutes of speech per individual was undertaken. This is displayed in table 4.19, with the results of the t-tests in EXCEL recorded. (For reasons of space and presentation, the tables for the rest of the social variables are given in appendix 7; table 4.19 is given here as an example). Note: t-test is the significance level for each set at $p < 0.05$.

Table 4. 19: Pairwise t-test for lexical borrowing by age groups

Age groups	Young and middle	Middle and old	Old and young
t-test	0.795987	0.011475	0.010364

The t-tests show that there is no significant difference between the young and middle groups in patterns of borrowing ($p = 0.796$), but that the old group is significantly different from the middle ($p = 0.012$) and young groups ($p = 0.010$).

Table 4. 20: Age groups in relation to code-switching

Code-switching	Switched from	Young	Middle	Old	Total
Single words	English	15	2	2	19
	Swahili	31	13	13	57
Phrasal switches	English	22	8	3	33
	Swahili	73	91	22	186
Sentences	English	-	-	-	
	Swahili	32	26	1	59
Tags	-	-	2	1	3
Total	**English**	**37**	**12**	**6**	**55**
	Swahili	**136**	**130**	**36**	**302**
Grand total		**173**	**142**	**42**	**357**

Concerning code-switching, youngsters and middle-aged speakers code-switch more frequently than old speakers. While there were 32 and 26 switched sentences to Swahili in the young and middle age group respectively, there was only 1 sentence from the old-aged speakers. There were only 25 phrasal switches from the old speakers, of which 3 were English forms and 22 were from Swahili. Young speakers switched up to 95 phrases, of which 23 were switches to English and 72 to Swahili, while middle-age speakers switched 99 phrases, of which 92 were to Swahili forms and 7 were switches to English. Single word switches occurred more frequently among the young people than the middle-aged and old people. Out of 72 single word switches, 43 were from young speakers, and 13 were switched by the middle–aged group, while 16 were switched by old-aged speakers. Of these, 13 switches to English single words are by young, 3 are by middle-aged and 3 by old speakers. Table 4.20 indicates that there are fewer English

switchings compared to Swahili ones, and it can be concluded that young speakers are leading in code-switching. Furthermore, switches from English are common in this group compared to middle and old-aged groups. However, since there is uneven distribution of speakers among age groups uneven, a pairwise t-test was conducted through EXCEL to evaluate how significant the difference is between age groups. As with lexical borrowing above, the t-test involved dividing the number of tokens per speaker by the length of interview in minutes times five. The results are as in table 4.21 below.

Table 4. 21: Pairwise t-test for code switching by age groups

Age groups	Chasu to Swahili	Chasu to English
Young & middle	0.777537	0.219032
Middle & old	0.002667	0.639917
Young & old	0.014083	0.203123

Concerning switching to Swahili, the t-test shows that there is a significant difference between young and old (p=0.014083) as well as old and middle (p= 0.002667) but not between young and middle (p=0.777537). On the other hand, with switching to English the difference between age groups is insignificant, as table 4.21 indicates. This implies that the effects of language contact are higher among young and middle-age speakers, especially for Chasu and Swahili.

4.3.2 Gender

Table 4. 22: Gender in relation to lexical borrowing.

Male n=31 (54.3%)			**Female n=26 (45.6%)**		
Lexical borrowing	**Core borrowing**	**Cultural borrowing**	**Core borrowing**	**Cultural borrowing**	**Total**
Verbs	59	36	80	22	
Nouns	130	351	99	240	
Adverbs	12	-	8	-	
Adjectives	6	-	-	-	
Conjunctions	1	-	5	-	
Pronouns	-	-	-	-	
Prepositions	-	-	-	-	
Numerals	400		231		
Total	**608**	**387**	**423**	**262**	**1680**
Grand total	**995**		**685**		
t-test	**male vs. female = 0.357742**				

Table 4.22 illustrates the rate of lexical borrowing among males and females. The data shows that men used 995 of the total lexical words, of which 608 are core and 387 are cultural borrowed forms while women used 685, of which 423 are core and 262 are cultural borrowings. If we excluded borrowed numeral forms core and cultural borrowings would respectively be 208:387 for male and 192:262 for female, implying that cultural borrowings in fact outnumber core borrowings in both sexes. Generally, men appear to borrow more than women. However, since the number of speakers is different in that we have fewer females than males, the t-test, as described above, was used to appraise these quantities for the full totals, including numerals. Though table 4.22 shows that men use more loan words than women, the t-test in the last column indicates that the difference between men and women is 0.357742, which is not significant. More information concerning the rest of the other variables is in appendix 8.

Table 4. 23: Gender in relation to code-switching

Types of code-switching	Switched from	Male n=31 (54%)	Female n=26 (46%)	Total
Single-words	English	4	15	19
	Swahili	19	38	57
Phrasal switches	English	12	21	33
	Swahili	124	62	186
Sentences	English	-	-	
	Swahili	32	27	59
Tags		3		3
Total		**194**	**163**	**357**
	English	19	36	55
	Swahili	175	127	302
t-test	**For Swahili switching**		**0.515858**	
t-test	**For English switching**		**0.168968**	

In code-switching, men's code-switches totalled 194 compared to women, who have 163 occurrences of switches. Thirty-two sentences were switched by men, while 27 were from women. Men switched up to 136 phrases, of which 12 were English forms while 83 phrasal switches came from women, of which 21 were from English and 62 were Swahili forms. In terms of single-word switches, men used 23 forms, of which 4 were English forms, while women had 49 forms of which 15 are from English. In this context, women use more single word switches and phrases from English than men, who use more forms from Swahili. However the t-test for between male and female Swahili switching is 0.515858 and for English switching is 0.168968, implying that there is no significant difference in switching.

4.3.3 Social class

As mentioned before, social class in this study is made up of two groups. The first is the middle class, composed of professionals and businessmen and women, and the second is the lower class, made up of pastoralists and peasants. In chapter five it will be explained why these four occupational groups were collapsed into two major groups, middle and lower class. Tables 4.24 and 4.25 present the data on lexical borrowing and code-switching by social class.

Table 4. 24: Social class in relation to lexical borrowing

Middle class n=29 speakers			**Lower class n=28 speakers**		
Lexical borrowings	**Core borrowings**	**Cultural borrowings**	**Core borrowings**	**Cultural borrowings**	**Total**
Verbs	83	46	57	12	
Nouns	115	344	121	239	
Adverbs	19	-	1	-	
Adjectives	1	-	5	-	
Conjunctions	5	-	1	-	
Pronouns	-	-	-	-	
Prepositions	-	-	-	-	
Numerals	377	-	254	-	
Total	**600**	**390**	**439**	**251**	**1680**
Grand total	**990**		**690**		
t-test	**0.000138**				

Table 4.24 relates lexical borrowing and social class whereby, out of 1680 borrowed words, 990 were used by middle class speakers, while 790 were used by lower class speakers. There is more frequent borrowing (of both core and cultural terms) in the middle class speakers than in the other group studied. Likewise, a t-test appraising the significance of the difference between middle and lower class shows (in the last column of table 4.24) that p= 0.000138 – a highly significant difference. For more information concerning the t-test in relation to social class, see appendix 10.

Table 4. 25: Social class in relations to code-switching

Code-switching		Middle class	Lower class	Total
Single words	English	15	4	**19**
	Swahili	36	21	**57**
Phrasal switches	English	32	1	**33**
	Swahili	119	67	**186**
Sentences	English	-	-	
	Swahili	53	6	**59**
Tags		3	-	**3**
Total	**English**	**50**	**5**	**55**
	Swahili	**208**	**94**	**302**
Grand Total		**258**	**99**	**357**
t-test	**Switching to Swahili**	**4.07691E-05**		
t-test	**Switching to English**	**0.003079**		

Concerning code-switching and social class, table 4.25 shows that from a total number of 357 switches, 99 forms were switched by the lower class, while 258 forms by the middle class speakers. While middle class speakers switched 51 single word forms, of which 15 were from English, lower class speakers switched 25 single words, of which only 4 were from English. Whilst the middle class speakers switched up to 151 phrasal forms, of which 32 were from English, 67 phrasal forms were switched by the lower class speakers, of which only 1 was from English. At the sentential level, the middle class score is 53, as against 6 from the lower class. Generally, middle class speakers code-switch more frequently, but also switch to English more frequently than their lower class counterparts. The significance of the difference in code-switching by the lower and middle classes was assessed through pairwise t-tests. The t-test for the difference in switching to Swahili by the middle and lower classes is 4.07691E-05 or (0.0000407691) and 0.003037 for Chasu to English switchings. The t-tests prove that the higher the speaker's level in the social hierarchy, the greater his/her chances of using borrowed words and switching codes in Swahili and English, and the difference is highly significant.

4.3.4 Education levels

Individuals acquire their first language during childhood. However if education is attained through a second language as a medium of instruction, it is evident that these individuals will end up being bilinguals. In addition, if the language used in formal social, political and economic activities is different from the first language, people are forced or at least motivated to learn that language for success in the society. This is evident in Tanzania in general, and in the rural Chasu community in particular. People learn a new language because it is a language of instruction in schools or colleges, or it is associated with significant opportunities. Tables 4.26 and 4.27 indicate how education level is related to lexical borrowing and code-switching. It is explained in chapter 5 why the social factor of education level was collapsed to "less educated" and "highly educated" speakers.

Table 4. 26: Education level in relation to lexical borrowing

	Highly educated n=20 (35%)		**Less educated n=37 (65%)**	
Lexical categories	**Core Borrowing**	**Cultural borrowing**	**Core borrowing**	**Cultural borrowing**
Verbs	61	44	79	14
Nouns	70	263	166	321
Adverbs	19	-	1	-
Adjectives	1	-	5	-
Conjunctions	5	-	1	-
Pronouns	-	-	-	-
Prepositions	-	-	-	-
Numbers	249		382	-
Total	**405**	**306**	**634**	**335**
Grand total	**711**		**969**	
t-test	**0.001214**			

Concerning lexical borrowing and level of education, the raw data indicates that less educated speakers used 969 borrowed words. On the other hand, the highly educated speakers used 711 borrowed lexical items as table 4.26 indicates. It is worth mentioning that this social group contains an asymmetrical distribution of informants, as a total of 57, 37 (65%) are less educated and 20 (35%) are highly educated. This disparity represents the population distribution in Chasu rural society, where highly educated individuals are scarcer than less educated ones. In order to appraise these differences, a pairwise t-test was perfomed in which p = 0.001214, as the last column of table 4.26 indicates. The difference between the highly and the less educated in using the borrowed words is significant.

Table 4. 27: Education levels in relation to code-switching

Code switching	**Switched from**	**Highly educated**	**Less educated**	**Total**
Single words	English	15	4	**19**
	Swahili	27	30	**57**
Phrasal switches	English	32	1	**33**
	Swahili	87	99	**186**
Sentences	English	-	-	**-**
	Swahili	29	30	**59**
Tags		3	-	**3**
Total	**English**	**50**	**5**	**55**
	Swahili	**143**	**159**	**302**
Grand total		**193**	**164**	
t-test	**Chasu to Swahili**		**0.002581**	
t-test	**Chasu to English**		**4.92124E-06**	

Similarly, table 4.27, which relates code-switching with educational attainment, shows that the highly educated speakers switch more frequently than the less educated ones. The highly educated speakers switch into English more frequently than the less educated ones. The highly educated group used 38 single word switches, of which 15

were from English, while the less educated have 34 single word switches of which only 4 were from English. While 120 of the phrasal forms were switched by highly educated people, of which 32 were from English, the less educated speakers switched 100 forms, of which only 1 is from English. The highly educated group switched 23 times at the sentential level, while the less educated switched 30 times. All 3 tags are from the highly educated speakers. The pairwise t-test for the difference in code-switching by the highly educated and the less educated speakers is p= 0.002581 for Chasu to Swahili switching and p = 4.92124E-06 for Chasu to English switching. (Appendix 9 gives more information concerning the t-test in educational levels). Generally, the t-tests show that despite the unequal representation of the speakers between groups, the highly educated speakers used borrowed forms and code-switched more frequently than the less educated speakers, and the difference is highly significant, especially in Chasu to English switches.

4.4 Discussion

From these social factors, it is evident that there are groups of individuals in multilingual Chasu society who borrow or code-switch more frequently, such as young people, highly educated and middle class individuals. Gender is not a significant factor in regulating either lexical borrowing or code-switching. As mentioned before, the socio-economic dominance and the status of the language used by the dominant group, and the degree and duration of contact may determine the scope of lexical borrowing and code-switching more frequently than other factors.

Concerning different age groups, most of the young and middle-aged speakers are bilinguals, exposed intensively to Swahili in a variety of contexts, far more than older informants who have casual contact with the language due to the nature of their daily social and economic activities. As explained in chapters 1 and 3 young speakers engage in business which exposes them to using Swahili more frequently as they travel far from home. Young speakers' interactions and social identities are forged by spending time on urban public entertainment like playing football and watching it on television, and watching movies and music on television etc. Thus their attitude towards languages is

highly focused on Swahili and English. Swahili is associated with nationalism, power, modernity and dynamic life (Mekacha 1993), while English is connected to new culture, entertainment, highly paying jobs and internationalism. This is like code-switching in urban Nairobi, where "by speaking some English, speakers associate themselves with the education, authority, and sophistication which this international language signifies…" (Myers-Scotton 1993b:12). This study is also similar to the patterns of code-switching between different creole varieties and European languages (Dutch and French) among the multilinguals of the Eastern Maroon (hereafter EM) community of Suriname and French Guiana (Migge 2007:53-73). In EM community, Dutch and French are associated with (white people's) power and knowledge, formal education and access to job opportunities. Dutch and French in the EM community are variable, and determined by level of education, residence patterns and occupation. They are mostly acquired by youngsters and people below 40, who are exposed to education and urban life. Sranan Tongo is the urban Creole associated with working class people, especially men, but also in the traditional EM ideology is depicted as lacking respect and power (Migge 2002). Lesipeki Taki is a variety in this community used for respectful traditional speech, informal socio-cultural events and ceremonies, interaction with and among elders and traditional titled persons, parent-child talk and women's subsistence work and leisure activities. In EM community, young speakers employ frequent code–switching to negotiate social identity, and make strategic use of Sranan Tongo, which is associated with the urban world, a cash labour economy and a modern western type of lifestyle. They construct an identity of being non-traditional and sophisticated, while elders use Lesipeki Taki and avoid Sranan Tongo because the notions of powerful elderhood are linked to engagement with EM tradition. In this study, some old speakers are completely monolingual in Chasu. They use few borrowed words and rarely code-switch to Swahili. Their attitude toward language use promotes preservation of ethnic language and culture. In addition, there is little exposure to entertainment or related activities which may bring new vocabulary. The older generation engage in farming and pastoralism, which confine them to one locale. Few have attained high levels of education which may expose them to Swahili and English usage.

Concerning gender, we have established in chapter 1 that women in rural Chasu practise farming activities after completing their primary school. Most men and some women go to school and acquire higher education, which exposes them to using more than one language. Women may engage in small business like selling vegetables, fruit or fish from neighbouring areas, but they do not frequently travel to the city areas (unless they have children living in the city, whom they visit once or twice a year. They have to perform domestic chores including rearing children and serving their husbands. Observation shows that it is rare to find women watching football or movies on television in public locations. This is because most of those leisure activities are screened in the evening, when women are customarily expected to be at home. Nevertheless the data has shown that with regards to gender, the difference in code-switching and lexical borrowing between men and women is not significant. Irrespective of the sex differences, men and women code-switched and used borrowed forms in the same frequency.

Relating to the dominant group, Chasu society has individuals who have attained high education, some of whom are doing professional jobs and/or are in business. These individuals are commonly found in the middle class, and in most cases this group includes youngsters and few middle-aged speakers. Though educated people might be few in number according to the population distribution, they comprise a dominant group in terms of socioeconomic and political achievements. They use Swahili and sometimes English professionally, reflecting their adoption of Western culture. These are the source sof new words in Chasu. Other classes follow the lead of the professionals in adopting loan words.

As explained in chapter one, the use of Swahili is a national concern in Tanzania. ECLs are inadequately drawn into competing with the mainstream society of people using Swahili. At the same time, the mainstream group also borrows culture from the outside. This poses a great challenge to the language policy-makers concerning the use of ECLs, which are supposed to be resources to Swahili, as stipulated in the language policy. People are focused on new culture, and the economic upward mobility associated with English and Swahili, rather than the culture connected with ECLs. Instead of picking

up words from ECLs to strengthen Swahili, as the language policy asserts, Swahili borrows from English, while ECLs borrow from Swahili and sometimes English. Consequently, the ECLs are compromised.

4.5 Chapter conclusion

This chapter has attempted to examine the structural effects of the language contact between Swahili and Chasu, and contact between Swahili, Chasu and English, through lexical borrowing and code-switching. The data has shown that there are borrowed words from both Swahili and English in Chasu. The level of code-switching extends from single word lexemes and phrases to clauses and sentences. Code-switching involving Swahili and English is widespread among the few young, highly educated middle-class speakers. There are no switches from English at sentential level yet. These structural forms may occur due to high infusion of languages in contact, but it may also be a way of reflecting social values, especially if the languages in contact carry different social meaning and status value. When one language is less prestigious in terms of use and its associated culture, people will end up borrowing and code-switching to identify with the more prestigious language and culture. Haugen (1972:199) clarifies this by stating that 'words are often borrowed when they are felt to be prestigious or just novel…this is especially true if speakers feel inferior to the speakers of the other language. The loan words may cause native words to seem inadequate and gradually disused'. Thus, as a way of reflecting social values, the data has shown that women, youth, highly educated and middle-class speakers used borrowed words, and code-switch both in Swahili and English compared to the rest. Myers-Scotton (2006:212) comments that the tendency to borrow from English has been associated with its uses in the advancement in science and technology, and its being a language of modernity, but it appears that borrowings are in more domains than simply science and technology. This implies that if the level of contact between these languages is constantly intensified, we can anticipate more changes in the structure of Chasu.

CHAPTER FIVE
PHONOLOGICAL VARIATION IN SOUTHERN CHASU

5.0 Introduction

The previous chapter investigated sociolinguistic variation in the context of language contact, particularly lexical borrowing and code-switching. The chapter at hand investigates another aspect of variation, viz. phonological variation within Chasu itself. It undertakes a multivariate analysis of the phonological variation of (z) and (s). It also discusses how social variables control language variation in the phonological context. The presentation and discussion of the data is preceded by a description of the occurrence of these phonological variables in Southern Chasu and the neighbouring languages. This is an extended review of what was introduced in chapter one.

5.1 Variables (s) and (z) in Southern Chasu and neighbouring languages

As discussed in chapter one, Kotz (1909) and Mreta (1998) have suggested that the major difference between Northern and Southern Chasu occurs at the phonological and lexical levels. Kotz and Mreta indicate that the Southern dialect has voiceless alveolar fricative [s] and voiced alveolar fricative [z], corresponding to the Northern variety voiceless dental fricative [θ] and voiced dental fricative [ð]. Gweno, which is linguistically related to Northern Chasu, uses [ð] and [θ] interchangeably with [z] and [s] of the Southern dialect. Together with other examples given in chapter one, the following examples in table 5.1 demonstrate the difference between the Southern and Northern dialects in relation to these sounds. Data for the Southern Chasu in table 5.1 is based on the written Chasu literatures like Bibles and hymn books.

Table 5. 1: The occurrence of (s) and (z) in Southern and Northern Chasu

Southern dialect	Northern dialect	Gloss
i-**z**iso	i-ðiθo.	eye
m-go**s**i	m- ɣoθi	husband
i-**s**aŋga	i-θaŋga	earth
ku-**s**ara	ku-θara.	to walk fast
m-**s**i	m-θi.	day
ku-**sa**gura	ku-θaɣura	to select
sa	θa	like
si	θi	down
cha-**s**ia	cha-θia	is finished
ma-**s**ambi	ma-θambi	traditional skirt
ŋga**s**u	ŋgaθu	traditional youth initiations
mi-ta**s**o	mi-taθo	prayers /churches
mu-e**z**a	mu-eða	tall person
y-e**z**ie	y-eðie	which came
ku-**z**oka	ku-ðoka	to travel
a-ni-**z**ori-ra	a-ni-ðori-ra	he has bought something for me
va-**z**ima	βa-ðima	elders/parents
ku-**z**a	ku-ða	to come
i-**z**ina	i-ðina	name
ke**z**ia	keðia	greet

Source: Data for the Northern dialect is from Mreta (1998)

As explained earlier, Kimambo (1969) points out that at different periods of time, groups of people moved northwards from Usambara Mountain, looking for new pastures, while another group of people moved southward, from Taita, then Mwanga and Ugweno, to Gonja and later Mamba. In the pilot field work for this study, elders could narrate the historical background of the Vaasu, which supports Kimambo's ideas. Elders report that *Vashamba - Vambuzii*, *Vamtewe*, *Vakirindi*, and later on *Vambughu*, moved from Usambara to the southern parts of Same district, but mostly dwelt in the southern part of Mamba as herdsmen. The Vabwambo moved from Taita through Mwanga and Ugweno, then Gonja, and later Mamba, looking for land for cultivation. They were later followed by the Vamjema, who came to be local rulers, *Vamfumwa*. The migrants encountered the

inhabitants, the *Vaasi* (meaning "cursed people"), who were herdsmen and known to be physically very short. The Vaasi were defeated by their adversaries in wars to the point of near-total extinction. However, it is difficult to establish accurately the period of migration and the group which started to migrate to Vuasu area. Migrants from various directions have mingled to the extent that it is difficult to differentiate them unless they identify themselves through their clan names or through their culture, which differ according to the directions they moved from. It was explained in chapter one that since most of the southern Vaasu are from the south, i.e. the Usambara mountains, Southern Chasu is very close to the Shambala language as well as Mbugu (especially the Normal Mbugu dialect) in terms of its phonological and lexical attributes (Kagaya 1989, Mous 2003). Together with examples given in chapter one, the following examples in table 5.2 and 5.3 show how these related dialects maintain [z] and [s] in cognate lexis.

Table 5. 2: Lexical and phonological relation between Shambala and Southern Chasu

Southern Chasu	Shambala	Gloss
sakame	**s**akame	blood
lu**s**a**z**i	u**s**a**z**i	bed
lu**s**oyoyo	u**s**oyoyo	to be worried
ma**s**a	ma**s**a	fault
siŋgano	**s**iŋgano	needle
sona	**s**ona	lick
i**z**i**s**o	**z**isho	eye
ki-te**z**u	nte**zu**	basket
ku-**z**a	ku-i**z**a	to come
m-**z**iii	m-**z**iɣi	medicinal herbs
ma**z**i	ma**z**i	water
mba**z**i	mba**z**i	sympathy

Source: Besha (1989)

Table 5. 3: Lexical and phonological relations between Normal Mbugu and Southern Chasu

Southern Chasu	Normal Mbugu	Gloss
nta**s**a	nhta**s**a	barren animal
i**s**ibo	i**s**ibo	thick stick used by old people to support walking
m**s**avi	m**s**avi	sorcerer
m**s**aŋgo	m**s**aŋgo	earthworm
suiža	**s**uija	forbid
sakaža	**s**akaja	wear out
sano	**s**ano	five
i**z**iaka	i**z**iaka	quiver
i**z**uva	i**z**uva	sun/ day time
i**z**oka	i**z**oka	axe
m**z**ii	m**zi**gi	medicinal herbs
i**z**uŋgo	i**z**uŋgo	small bush made by calabash plant
ziŋgi**z**a	**z**iŋgi**z**a	shiver, shake
i**z**ana	i**z**ana	day before yesterday

Source: Kagaya (1989)

Currently, in Southern Chasu, variables (s) and (z) are retained in written texts, as column 1 in tables 5.1, 5.2 and 5.3 above illustrate. However, in the informal speech (s) occurs as either variant [s] or [θ] while (z) is realised as either variant [z] or [ð], as table 5.4 illustrates.

Table 5. 4: Sounds [s] or [θ] and [z] or [ð] in Southern Chasu

Southern Chasu		Gloss
[s] and [z]	[θ] and [ð]	
ki**s**aka	ki**θ**ka	small bush
m**s**au	m**θ**au	soil
i**s**aŋza	i**θ**aŋza	morning
ka**s**io	ka**θ**io	baby bird
i**z**au	i**ð**au	wing (of a bird)
m**z**uka	m**ð**uka	ghost
lu**z**ii	lu**ð**ii	rope
lu**z**oa	lu**ð**oa	feather

It is consequently hypothesized that [ð] and [θ] may be found in Southern Chasu as a result of contact with the people who moved from the north through Mwanga and Ugweno or directly from the Northern Chasu dialect. It is difficult, also, to establish a firm geographical demarcation between the Southern and Northern Chasu dialects, due to the intermarriage and mingling of people from these two geographical areas. As a result, scholars like Kagaya (1989) have written descriptions of Southern Chasu based on the Mbaga and Gonja sub-dialects. However, Kagaya's description does not adequately describe Southern Chasu, as his book includes some Northern Chasu phonological features and vocabulary items that do not occur in the South.

Swahili has [ð] and [θ], particularly in words borrowed from Arabic, as in the examples in table 5.5. These words are hardly used in Chasu (except when one is code-switching to Swahili, as mentioned in chapter four, especially with numerical expressions), because some of them retain their equivalents in Chasu.

Table 5. 5: Sounds [ð] and [θ] in words switched to Swahili

Swahili	Gloss
θamani/ θamini	value (noun)/verb
θeluji	snow
θelaθini	thirty
θemanini,	eighty
θumuni	fifty cents
ðihaka,	insult
mðamini	sponsor
ðambi	sin
ðuru,	harm
ðumuni,	objective/aim

In the word structure of the Southern Chasu dialect, the [z] (which may sometimes occur as [ð]) is found in different word categories, such as verbs, nouns, adjectives and conjunctions, used in different environments, as in the following examples;

1. Initial syllable
   (a) **z**ora- 'buy'
   (b) **z**ura- 'select'
   (c) **z**iaho- 'although'
   (d) **z**aŋga- 'dating (for courtship)'

2. Middle syllable
   (a) m**z**oro- 'a servant'
   (b) i**z**ina- 'a name'
   (c) ki**za**ŋgo-'forehead'
   (d) m**z**umbu- 'ants'

3. Final syllable
   (a) kia**z**e- 'appointment'

(b) ma**z**i- ‘water’

(c) ku-**z**a- ‘to come’

(d) -e**z**a- ‘long’

(e) u-ti-**z**e – you-say- what/how –‘what or how do you say’

(f) ka**z**a- ‘praise (given only to deity)’

Except for (c) in example 1, which is a conjunction, (a), (b) and (d) are verbs. When [z] is attached to a verb in initial position, it denotes imperative voice, i.e. for giving orders or commands. [z] may also occur as the first element of a verb root, taking a word-medial position if there is an infinitive marker **ku-,** or subject pronouns and tense marker. With nouns, [z] may occur in initial syllable and is normally pre-nasalized to form a syllabic unit, as in words like ***nzi****a-* ‘way’, ***nza****ta-* ‘a stick’, ***nzo****ta-* ‘famine catastrophe*’*. As mentioned in chapter one, [z] is the only fricative sound among other Chasu consonants which is pre-nasalized when it occupies the initial position in a, Chasu noun. According to the data, [z] occurs frequently in the middle position in Chasu nouns as in example 2, or, rarely, with verbs that have a prefix. In example 3, [z] usually occurs in the final syllable in nouns, as in 3(a) and (b). However, exceptions arise for some verbs. In 3(c) we have the verb ***ku****-za-* ‘to come’, this is one of the monosyllabic verbs in Chasu, which are always prefixed with the infinitive aspect **ku**- to accommodate a stress position which should be in the penultimate syllable. In 3(d), there is *–eza*, which is a root of the adjective ‘long’. The adjective always has to be prefixed with subject agreements from different noun classes, such as *mu-/va-eza;* for animate (MU/VA), *nd-eza* for nasal noun class (N/N) etc. Example 3(e) *–ze* is a suffix, which is attached to a verb (always occurs as the final syllable) meaning ‘how’ (for condition and for the instrument), as in

4. (a) u- tonga -ze,
   you- go- how
   ‘How do you go’ (i.e. by what means will you travel)

(b) a-kaa-ze
   she –is –how
   ‘How is she?’ (i.e. for health condition not greetings).

Example 3(f) is for the verb 'to praise'. I have mentioned before that we do not get verbs with [z] in the final syllable, but *kaza* 'to praise' is an exception. With much more data we might come across a few more such verbs. Generally in Southern Chasu, we expect [z] in the initial syllable in the verbs, the middle syllable with nouns and the final syllable with nouns as well as a few other word categories.

On the other hand, [s] is found in different grammatical categories such as verbs, nouns, adjectives, conjunctions etc. It can occur in word initial, middle or final syllable, despite the type of word category. Unlike [z] above, [s] is attached to nouns as initial, middle and final syllable. We rarely find verbs with [s] in the final syllable; at least in the data we have, none was found. In a sentence structure it can also emerge as a monosyllabic grammatical category, as in *sa/se-* 'similar to', or as the verb 'to be' negative *si-* 'is not.'

5.2 Data Results

Whereas in the previous chapter on language contact the effect of each external (social) factor was analysed separately, this chapter tries to analyse, via more sophisticated statistical means, which factors are more influential than others, taking into account the fact that the social factors are not independent; for example, the same person could be 'young' and 'educated' to 'middle class level' and 'female'. It was hoped that VARBRUL, the standard statistical tool (Paolillo 2002, Tagliamonte 2006), would tease out these effects and indicate which social factors were more influential. In order to do this, I undertook the standard coding procedure associated with VARBRUL analysis.

After the identification of the dependent variables with their variants, the application value was assigned to each variant, where (0) represents [s] while (1) is [θ] for the first run, and (0) represents [z] and (1) stands for [ð] for the second run. The next step was to code the independent variables into the token file for the Goldvarb program, to estimate the probability or the values for each contextual factor specified.

The coding symbols were:

0	**i**	**b**	**n**	**f**	**y**	**1**	**1**	**z**	**w**
1	**m**	**f**	**y**	**m**	**m**	**2**	**2**	**p**	**c**
	f				**o**	**3**		**h**	
						4		**t**	

Where, both letters and digits were used to represent the factor groups as follows:

Factor group 1: Variants
0: variant [s] or [z]
1: variant [θ] or [ð]

Factor group 2: Syllable position
i: initial syllable
m: middle syllable
f: final syllable

Factor group 3: Vowel environment
b: back vowel
f: front vowel

Factor group 4: Status of lexical item
n: Chasu words
y: borrowings from Swahili

Factor group 5: Gender
f: female
m: male

Factor group 6: Age
y: young age group 21-40 years
m: middle age group 41-60 years
o: old age group 61+ years

Factor group 7: Social class (occupation) **1**: professionals (teachers, nurses, leaders)
2: businessmen and women
3: peasants
4: pastoralists

Factor group 8: Language use **1**: monolingual Chasu
2: bilingual in Swahili and Chasu

Factor group 9: Education level **z**: did not go to formal school at all
p: primary education
h: high secondary school
t: tertiary education

Factor group 10: Styles **w**: wordlist style
c: casual style

The token file contains the list of complete survey results for statistical analysis represented in the form of code. After the token file was created, the condition file was set up by the Varbrul software. The program also created an iteration file of independent variables known as a 'cell file'. This contains a summary of independent factors through counting. The Goldvarb software displayed the results of the proportion in terms of percentages for each factor, the number of contexts per cell, and the total number of contexts treated in the analysis. The result for the first run before recoding of the tokens is attached as appendix 3 for variable (z), and appendix 4 for variable (s).

Since we had few pastoralists and few individuals who did not attend any formal school, as mentioned in chapter three, I had to examine the distribution of data through cross-tabulation. Based on the cross-tabulation, exclusion and recoding of certain factor groups took place due to high interaction of factors. One factor group,"language use" (being monolingual of Chasu or bilingual of Chasu and Swahili), was eventually excluded due to insufficient variable representation from members of each group. For

instance, only 3 speakers were monolinguals, while the remaining 54 are bilinguals. In addition, two other social factors were recoded: social class and educational levels. Concerning social class, there was an unequal spread of pastoralists. This is due to the fact that possession of animals is entitled only to men in Chasu society. Furthermore, due to climatic changes, animal keeping is on the decline. Hence there was insufficient representation for pastoralists when it came to cross-tabulation with female gender. At the same time, only one pastoralist fell into the younger speakers group. Likewise, with educational levels, other groups had very few representatives; for instance, there were four speakers who did not pursue any kind of formal education, all above the age of 61, one female and three males. The data had only one speaker who acquired tertiary education aged over above 61. This meant that, when a multivariate analysis was performed with different independent variables, or the cross-tabulation with the dependent variables, a large number of empty cells, resulted. Thus, after recoding to avoid such empty cells I remained with two groups: (a) middle class (professionals and businessmen and businesswomen) recoded as **1,** as a replacement for **1**& **2**; and (b) lower class (pastoralists and peasants) recoded as **3** in place of **3** & **4**. In educational levels, I had the highly educated (tertiary and high school) recoded as **B** in place of **h** & **t,** and the less educated (primary school and no formal school at all) recoded as **A** in place of **p** & **z**. These regroupings make sense in terms of social criteria.

Based on the data after recoding (cf. Appendices 5&6 for detailed data after recoding), the following are the results of the first phase of the statistical analysis, which includes the percentages, number of tokens and frequency of occurrence in relation to variants [z] against [ð] and [s] over [θ]. Significance levels will not be discussed in this early section, but later on, in section 5.3, when considering the VARBRUL runs.

5.2.1 Results according to syllable position

Table 5.6 below presents the results of the VARBRUL run in respect of factor group for the syllabic position of variants [z] over [ð] as well as [s] against [θ]. Concerning the environment of the variants [z] and [ð] in the word structure, [z] occurs in

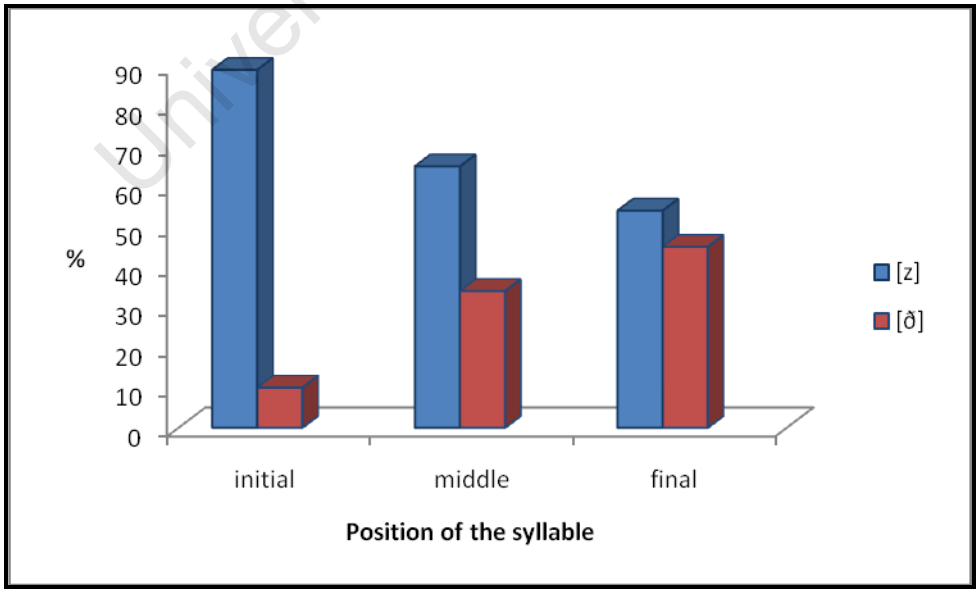
the initial, middle and final syllable positions with the percentage weight of 89%, 65% and 54% respectively. [Z] occupies over 50% of all positions, implying that variant [z] is dominant in the Southern Chasu word structure compared to variant [ð], as illustrated in graph 5.1 below.

Table 5. 6: Results according to syllable position

	[z]		[ð]					[s]		[θ]			
Factor group	**N**	**%**	**N**	**%**	**Total N**	**%**		**N**	**%**	**N**	**%**	**Total N**	**%**
Initial	128	89	15	10	143	12		493	78	135	21	628	37
Middle	407	65	210	33	605	54		335	58	241	41	576	33
Final	205	54	169	44	365	32		261	53	230	46	491	28
Total	**740**	**65**	**394**	**34**	**1134**			**1089**	**64**	**606**	**35**	**1695**	

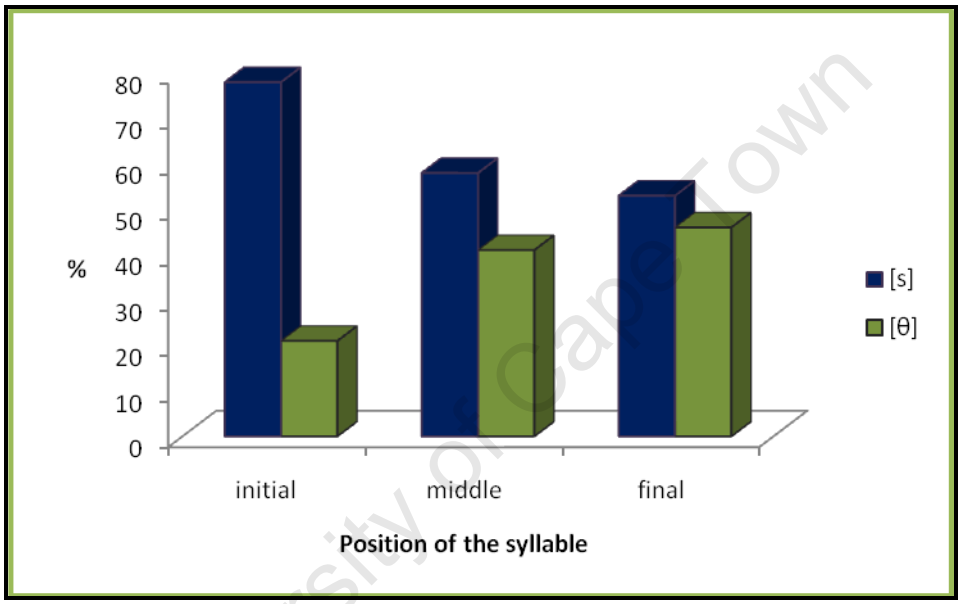
Despite the large number of tokens in the middle, as table 5.6 indicates, the percentage groups in graph 5.1 show vividly that, [z] is most dominant in the initial syllable position, then in the middle and lastly in the final syllable.

Graph 5. 1: Percentage of [z] and [ð] by syllable position



Turning to variants [s] and [θ] as syllables in the word structure of Southern Chasu, table 5.6 also indicates that the voiceless alveolar fricative [s] occurs at a percentage frequency of 78%, 58% and 53% in the initial, middle and final syllables respectively. The voiced alveolar fricative [θ] occurs 21% in the initial, 41% in the middle and 46% in the final syllable.

Graph 5.2 Percentage of [s] and [θ] by syllable position



The data in table 5.6 and graph 5.2 also exemplifies that, like [z], variant [s] occurs above 50% in all positions of the word structure, but most frequently in the initial position. Conversely, [θ] is the variant that occurs less than 50% in all positions. It has the reverse order, occurring most often in the final syllables, and least often in initial syllables.

5.2.2 Vowels following the variables

Table 5.7 presents the environment of the vowel following [z] and [ð] as well [s] over [θ], to examine what their effect is on the variables. Concerning the vowels following the variants [z] and [ð], table 5.7 indicates that, followed by back vowels, [z] is

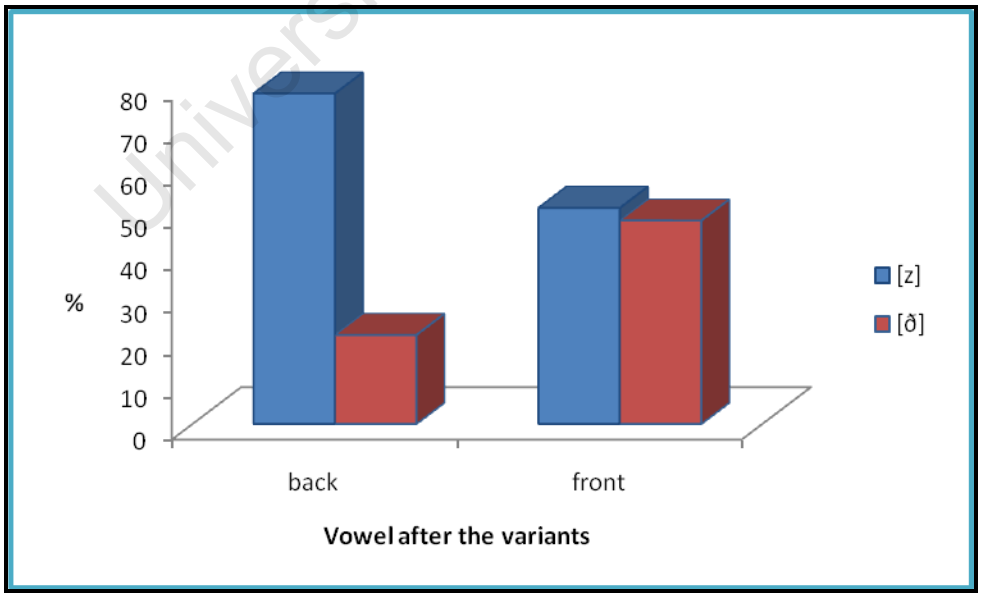
the variant 78% of the time and [ð] for 21%. On the other hand, 51% of front vowels occur after [z] while 48% follow [ð]. This implies that [ð] is the less common variant (below 50% each time) in all vowel environments, i.e. whether back or front.

Table 5.7: Vowels following the variants

	[z]		[ð]					[s]		[θ]			
Factor group	**N**	**%**	**N**	**%**	**Total N**	**%**		**N**	**%**	**N**	**%**	**Total N**	**%**
Back	460	78	126	21	586	51		461	60	306	39	767	45
Front	280	51	268	48	548	48		628	67	300	32	928	54
Total	**740**	**65**	**394**	**34**	**1134**			**1089**	**64**	**606**	**35**	**1695**	

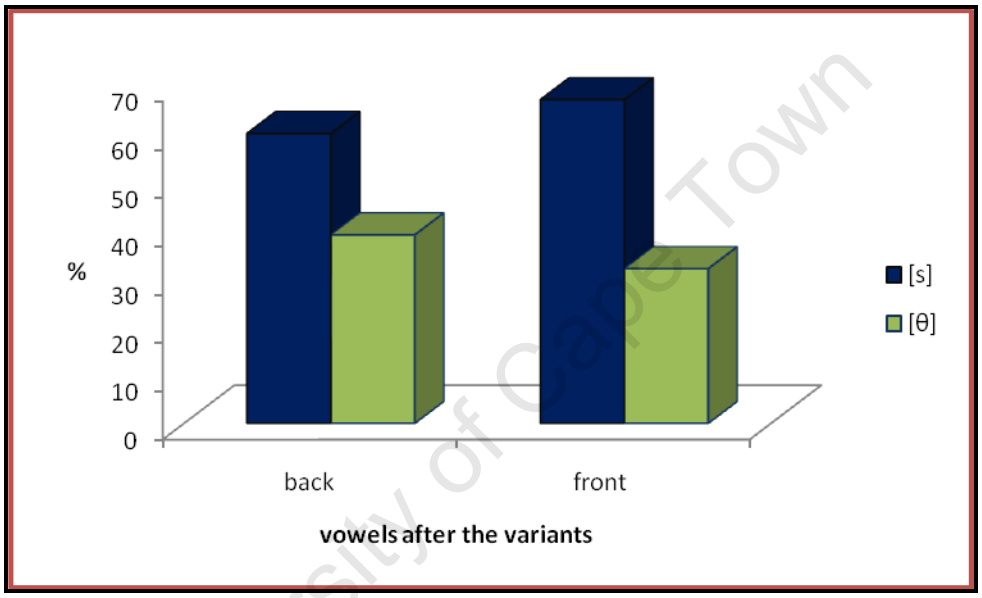
In graph 5.3, the illustration shows that the percentage difference for front vowels which follow variants [z] and [ð] is very small, while the difference for the back vowels is great. In the back vowel environment, it is the [z] variant that is more common.

Graph 5.3 Percentage of [z] and [ð] by vowel environment



Turning to variants [s] and [θ], table 5.7 also shows that on one hand, followed by front vowels; [s] is the variant 67% of the time and [θ] 39%. On the other hand, [s] occurs with back vowels 60% of the time, while the variant [θ] occurs 32% of the time. A high percentage of both back and front vowels occurs with variant [s], as graph 5.4 illustrates, while a smaller percentage favours [θ].

Graph 5. 4 Percentage of [s] against [θ] by vowel environment.



5.2.3 (s) and (z) in Swahili borrowings and native Chasu words

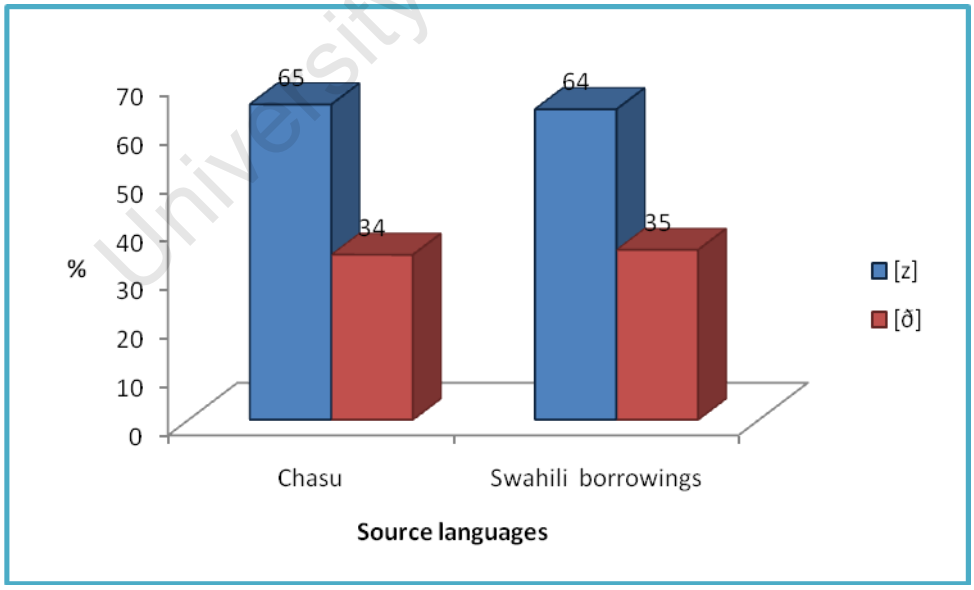
It was explained in chapters one and two that, as Swahili is a national language in Tanzania, simultaneous bilingualism and lexical borrowing is common attributes in numerous communities. In chapter four, I discussed the tendency to borrow words from Swahili into Chasu via new and intermediate cultures. Table 5.8 presents the percentage results in terms of the status of lexical items in relation to [z] and [ð**],** as well as [s] and [θ]. That is, whether there is a tendency for one variant to occur with Swahili borrowings more than native Chasu words or vice versa.

Table 5. 8: Status of lexical items in relation to variants

	[z]		[ð]					[s]		[θ]			
Factor group	**N**	**%**	**N**	**%**	**Total N**	**%**		**N**	**%**	**N**	**%**	**Total N**	**%**
Chasu	611	65	324	34	935	82		728	60	469	39	1197	70
Swahili	129	64	70	35	199	17		361	72	137	27	498	29
Total	**740**	**65**	**394**	**34**	**1134**			**1089**	**64**	**606**	**35**	**1695**	

Concerning [z] and [ð], the statistics in table 5.8 show the presence of 199 tokens from Swahili with the relevant variable, of which 64% favour [z]. On the other hand, 935 tokens are from Chasu, of which 65% favour [z]. The percentages for [ð] are thus 34% and 35% for Chasu native words and Swahili borrowed words respectively. Thus the [z] variant occurs in Swahili borrowings and Chasu native words in almost the same frequency.

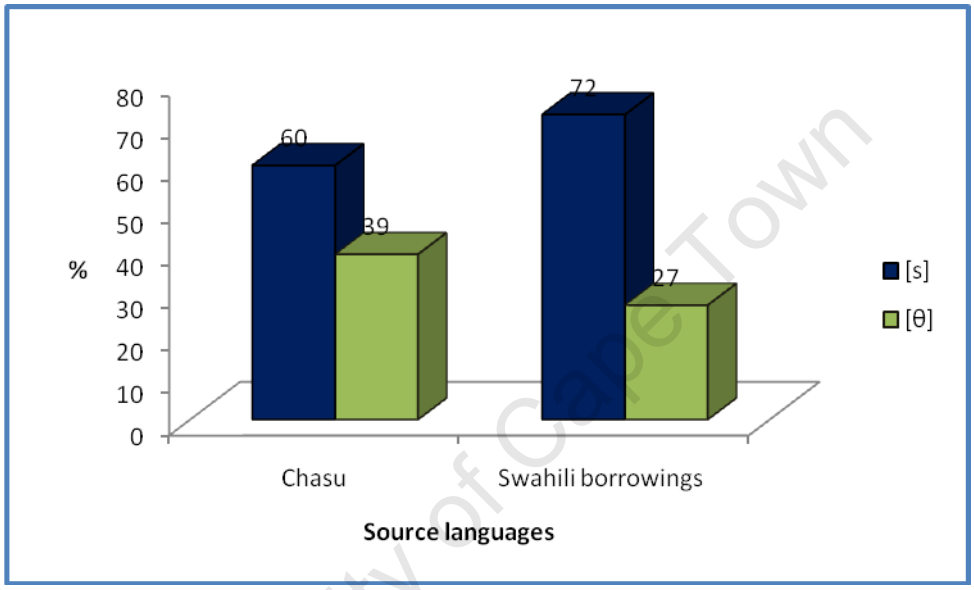
Graph 5. 5 Percentage of borrowings vs native words in relation to [z] and [ð]



Turning to [s] and [θ], table 5.8 above indicates that of all 1695 raw tokens obtained for variable (s), 498 are found in borrowed words from Swahili, while the

remaining 1197 tokens are from Chasu. Although the borrowed words from Swahili do not exceed the amount from Southern Chasu for this variable, the amount is sufficient to suggest the considerable impact of language contact on Chasu, as discussed in chapter four.

Graph 5. 6: Percentage of borrowings vs native words in relation to [s] and [θ]



Graph 5.6 above illustrates that 60% of native words from Chasu and 72% of the borrowed words from Swahili favour variant [s]; while the balance of borrowings 39% from Chasu and 27% from Swahili, occur with [θ]. Though [s] occurs frequently with words borrowed from Swahili and in native Chasu words, the percentage for the borrowed words is higher than for native words from Chasu. We examine statistical significances later in section 5.3.

5.2.4 Age groups

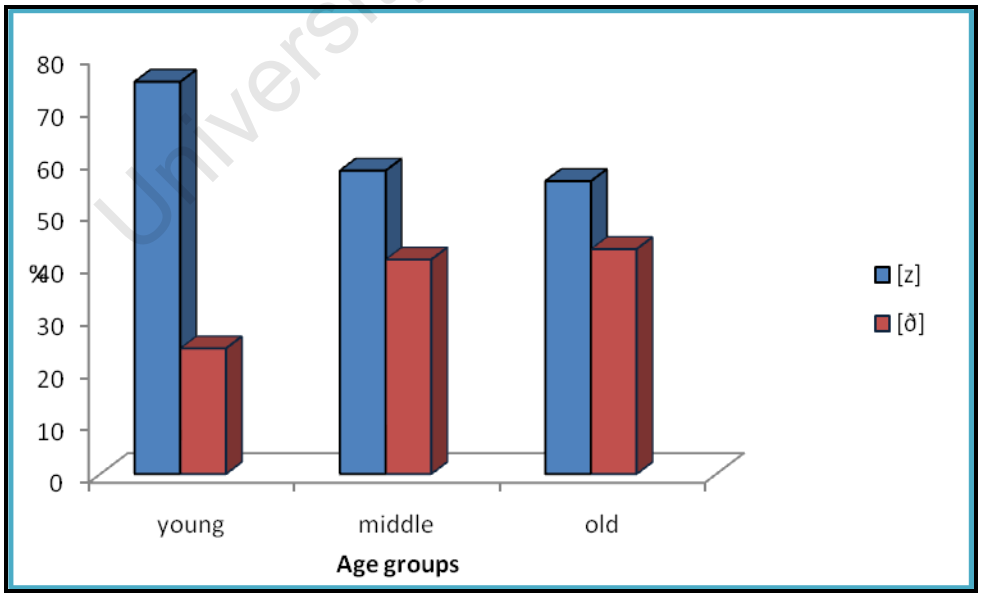
As mentioned before, this study contains three age groups: young (21-40 years), middle age (41-60) and old age speakers 61 years and older. Table 5.9 presents the results of percentage distribution of age groups in relation to [z] and [ð] as well as [s] and [θ].

Table 5. 9: Age groups in relation to variants

	[z]		[ð]					[s]		[θ]			
Factor group	**N**	**%**	**N**	**%**	**Total N**	**%**		**N**	**%**	**N**	**%**	**Total N**	**%**
Young	355	75	114	24	469	41		437	65	230	35	667	39
Middle	258	58	181	41	439	38		436	64	239	35	675	39
Old	127	56	99	43	226	19		216	61	137	38	353	20
Total	**740**	**65**	**394**	**34**	**1134**			**1089**	**64**	**606**	**35**	**1695**	

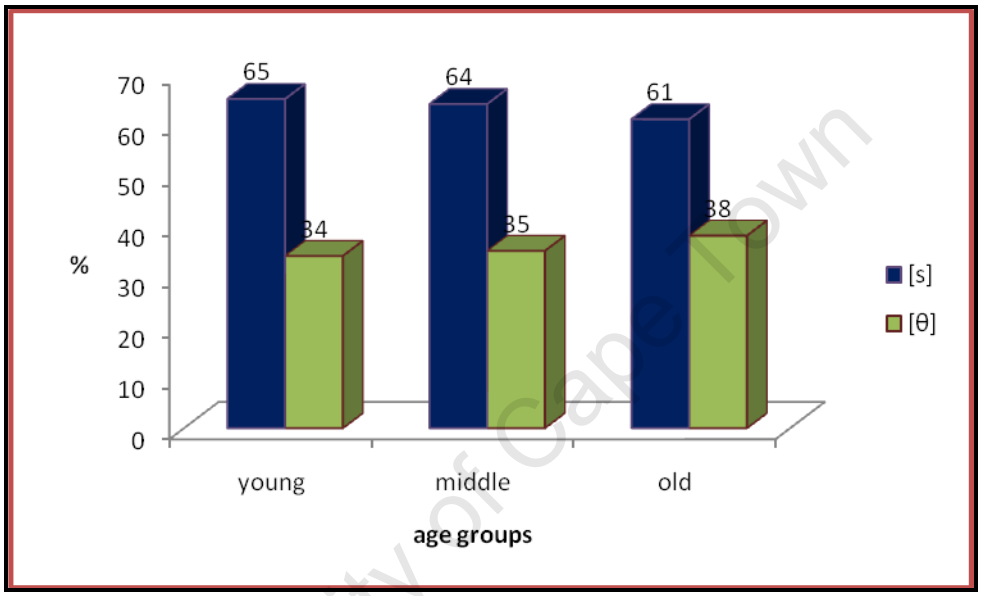
Starting with [z] and [ð], we see that all age groups use large proportions of [z] compared to [ð]. The favouring of [z] ranges from 56% for old age, followed by 58% for middle age to 75% for the younger age groups, as graph 5.7 illustrates. Though speakers from all age groups use variant [z] above 50%, the percentage for young speakers is considerably higher than for other speakers.

Graph 5. 7: Percentage distribution of [z] and [ð] among age groups



On the other hand , the figures in table 5.8 show that out of 667 raw tokens from young speakers, 65% had variant [s]; 64% of 675 tokens are from middle-aged speakers, while 61% of 216 came from old-aged speakers. The difference in favour of [s] against [θ] among the age groups is minute, as graph 5.8 illustrates below.

Graph 5. 8: Percentage distribution of [s] and [θ] among age groups



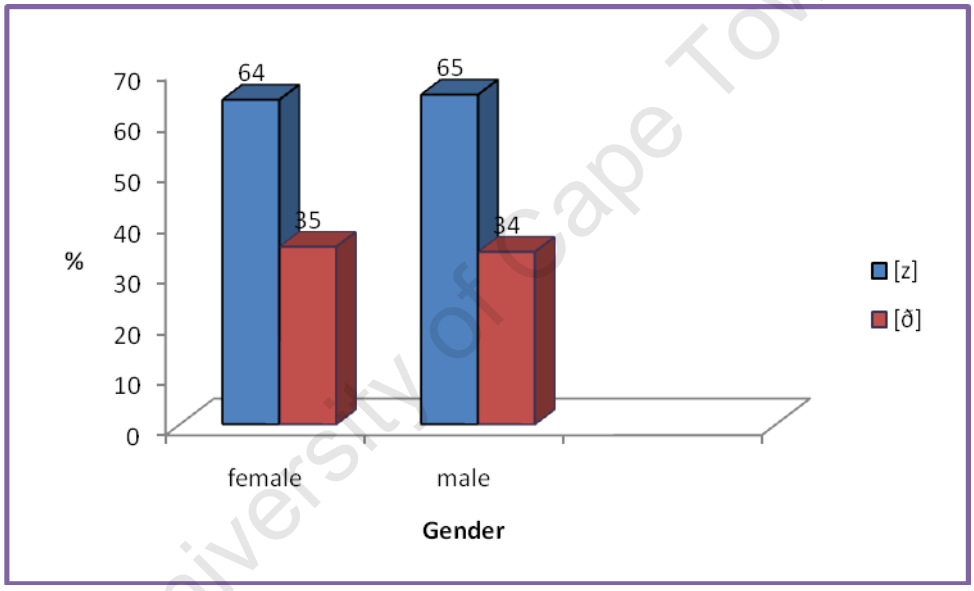
5.2.5 Gender

Concerning [z] and [ð], table 5.10 shows that the percentage distribution for [z] is 65% among males and 64% among females out of a total of 740 raw tokens. Judging by the percentage value, men and women speakers appear to use [z] against [ð] in almost equal frequency.

Table 5. 10: Gender in relation to the variants

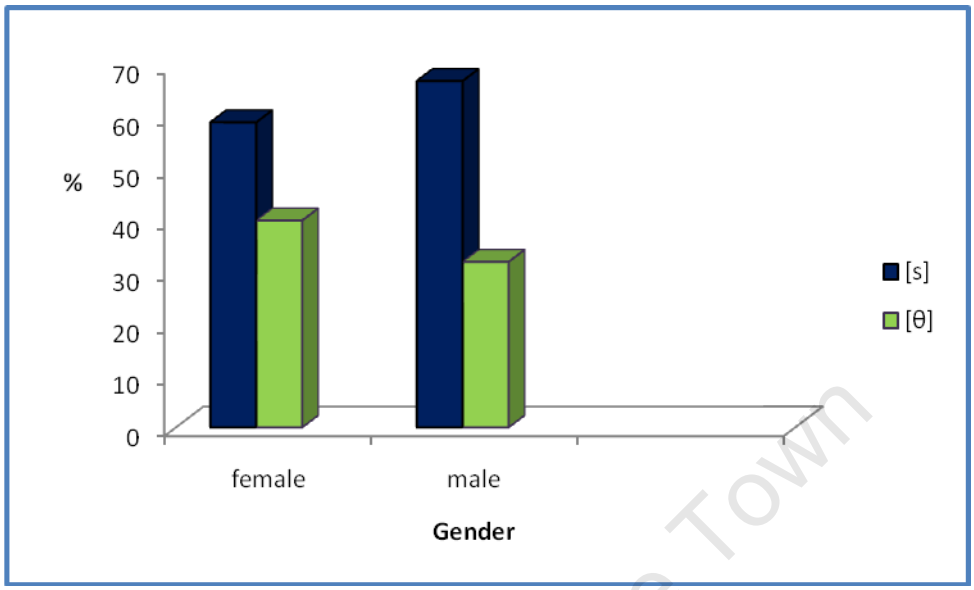
	[z]		[ð]					[s]		[θ]			
Factor group	**N**	**%**	**N**	**%**	**Total N**	**%**		**N**	**%**	**N**	**%**	**Total N**	**%**
Female	332	64	181	35	513	45		445	59	302	40	747	44
Male	408	65	213	33	621	54		644	67	304	32	948	55
Total	**740**	**65**	**394**	**34**	**1134**			**1089**	**64**	**606**	**35**	**1695**	

Graph 5. 9: Percentage distribution of [z] and [ð] by gender



Men's use of [s] is 67%, while women's use of the same variant is 59%, as graph 5.10 illustrates. Men's use of [θ] is 32% and women's is 40%. Though both men and women use [s] above 50%, the percentage value is higher for men.

Graph 5. 10: Percentage distribution of [s] and [θ] by gender



5.2.6 Educational levels

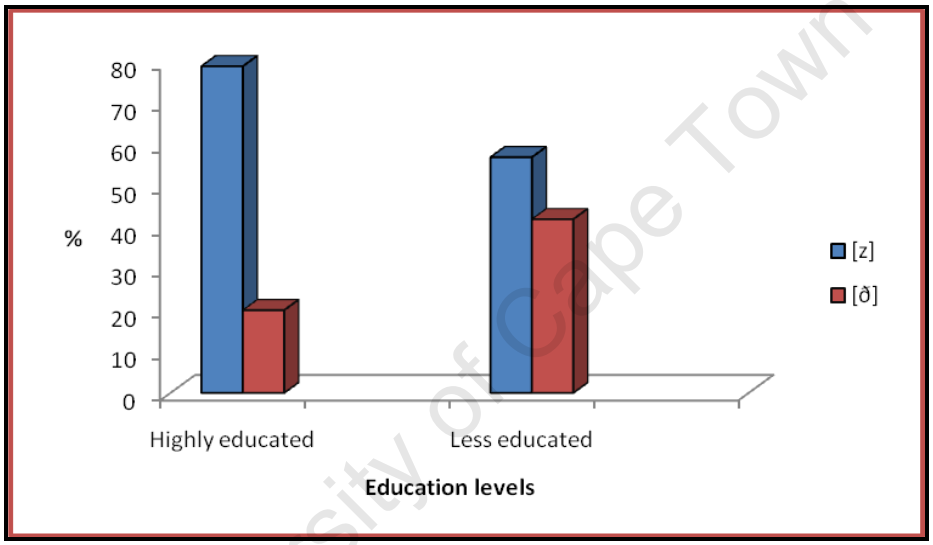
Table 5.11 presents the percentage results of education levels in relation to [z] against [ð] as well as for [s] over [θ]. Starting with variants [z] and [ð], the use of [z] by the highly educated speakers is 79%, and the less educated speakers 57%. The use of [ð] by highly educated speakers is 20%, while by the less educated it is 42%.

Table 5. 11: Education levels in relation to variants

	[z]		[ð]					[s]		[θ]			
Group factor	**N**	**%**	**N**	**%**	**Total N**	**%**		**N**	**%**	**N**	**%**	**Total N**	**%**
Highly educated	316	79	82	20	398	35		456	73	164	26	620	36
Less educated	424	57	312	42	736	64		633	58	442	41	1075	63
Total	**740**	**65**	**394**	**34**	**1134**			**1089**	**64**	**606**	**35**	**1695**	

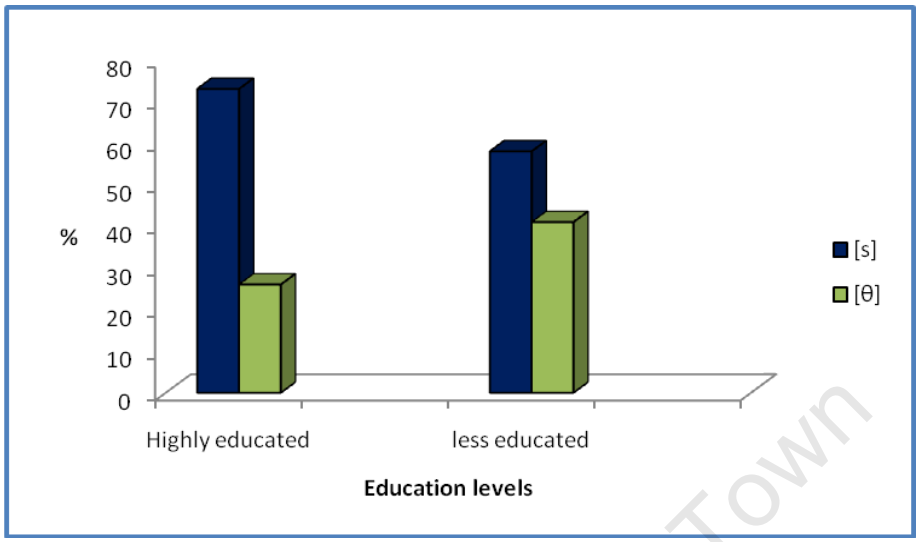
From the percentage value, speakers from both groups (less and highly educated) use [z] above the 50% level, though the difference is high, as graph 5.11 illustrates. The highly educated use [z] more frequently than the less educated.

Graph 5. 11: Percentage distribution of [z] and [ð] by education level



Turning to [s] and [θ], figures in table 5.11 show that the use of [s] by the highly educated speakers is 73% and by the less educated is 58%. The use of [θ] is 26% by the highly educated and 41% by the less educated. Speakers use [s] above 50% regardless of education level. However, the use of [s] by highly educated speakers is more frequent than by the less educated, as graph 5.12 illustrates below.

Graph 5. 12: Percentage distribution of [s] and [θ] by education level



5.2.7 Social class.

Table 5.12 illustrates the percentage value according to social class and its relationship on the frequency of occurrence of [z] and [s] over [ð] and [θ] respectively.

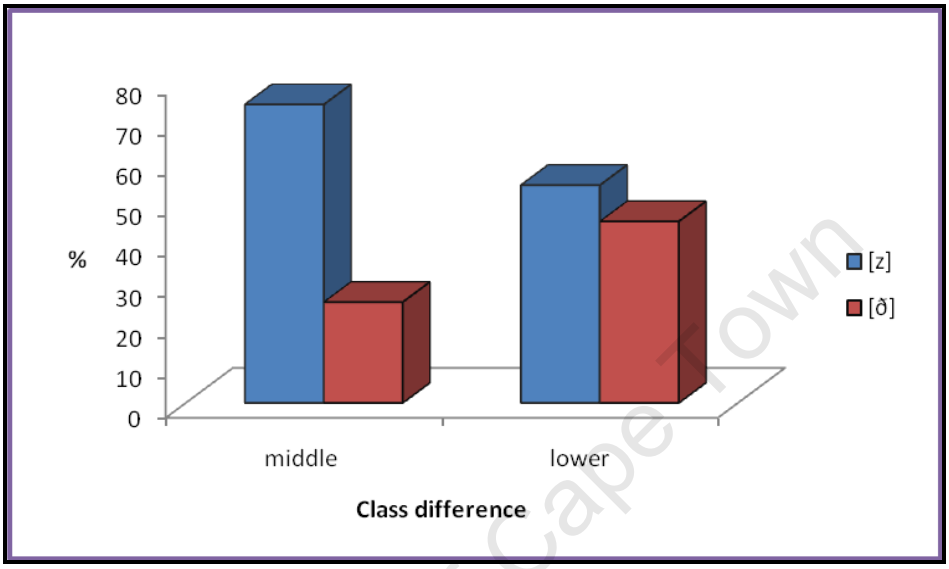
Table 5. 12: Percentage results according to social class

	[z]		[ð]					[s]		[θ]			
Factor group	**N**	**%**	**N**	**%**	**Total N**	**%**		**N**	**%**	**N**	**%**	**Total N**	**%**
Middle	448	74	154	25	602	53		635	69	279	30	914	53
Lower	292	54	240	45	532	46		454	58	327	41	781	46
Total	**740**	**65**	**394**	**34**	**1134**			**1089**	**64**	**606**	**35**	**1695**	

The middle-class speakers (professionals and business people) use of [z] is 74% out of 502 raw tokens, while the lower-class speakers' (peasants and pastoralists) use of [z] is 54% out of 532 tokens. Although the use of [z] by lower and middle-class is above

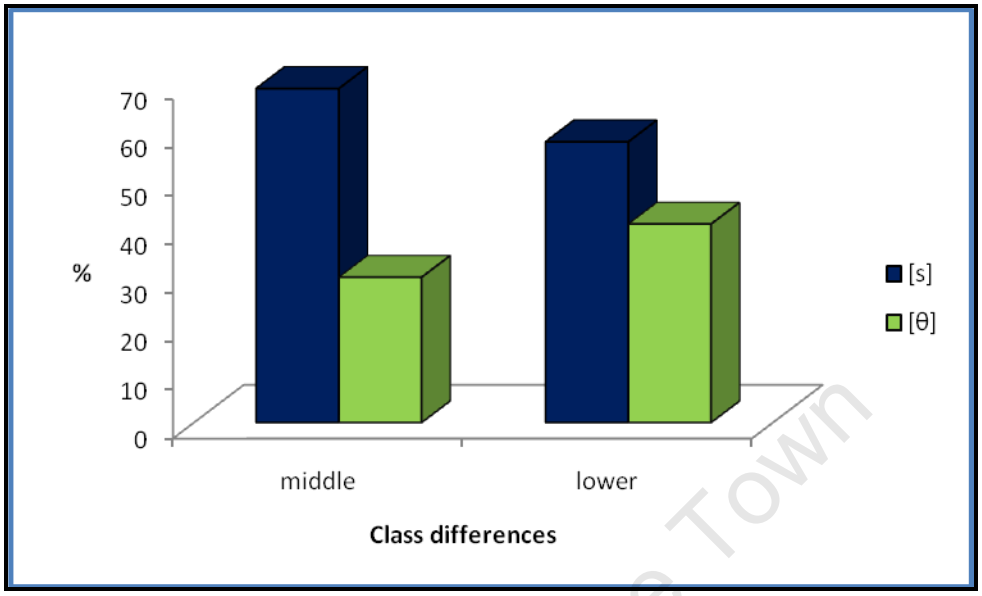
50%, middle-class speakers use [z] more frequently than [ð] compared to lower-class speakers, as graph 5.13 illustrates.

Graph 5. 13: Percentage distribution of [z] and [ð] by social class



Likewise, for the variable (s), middle-class speakers use of [s] is 69% compared to 58% of the lower-class speakers. The tendency indicates that, as for [z], the use of [s] is above 50% for all middle and lower-class speakers. However, middle -lass speakers use [s] more frequently than [θ] compared to the lower-class speakers, as graph 5.14 illustrates.

Graph 5. 14: Percentage distribution of [s] and [θ] by social class



5.2.8 Style

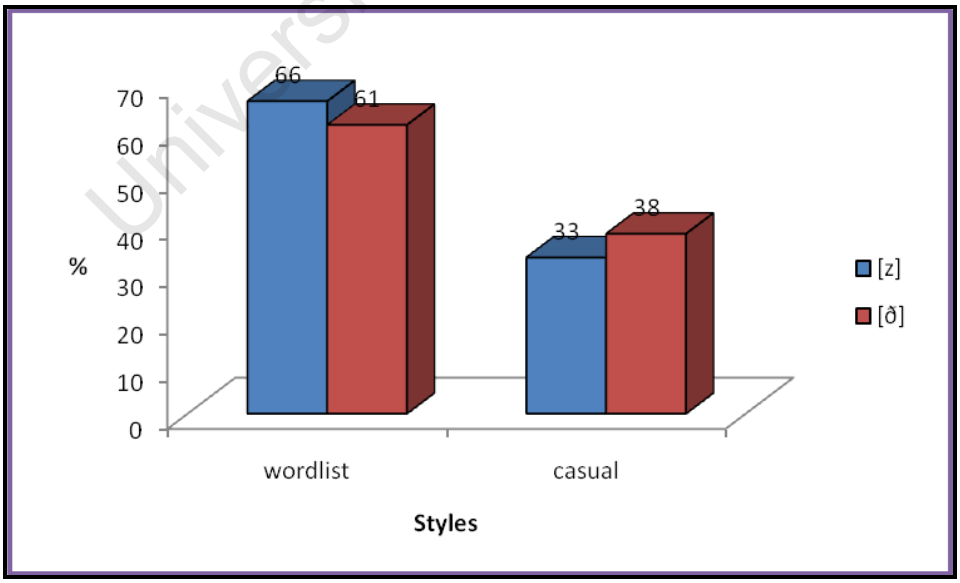
Following general sociolinguistic trends, I hypothesized that, on one hand, wordlist style is a way of accessing the style associated with formal and careful language, and with the standard form. On the other hand, conversation in the interview is a way of approximating casual or "natural" speech that includes non-standard features. Table 5.13 shows the percentage value for casual style and wordlist style in relation to variants [s] and [z] against [θ] and [ð] respectively.

Table 5. 13: Percentage results according to styles

	[z]		[ð]					[s]		[θ]			
Factor group	**N**	**%**	**N**	**%**	**Total N**	**%**		**N**	**%**	**N**	**%**	**Total N**	**%**
Wordlist	533	66	263	33	796	70		482	60	312	39	794	46
Casual	207	61	131	38	316	29		607	67	294	32	901	53
Total	**740**	**65**	**394**	**34**	**1134**			**1089**	**64**	**606**	**35**	**1695**	

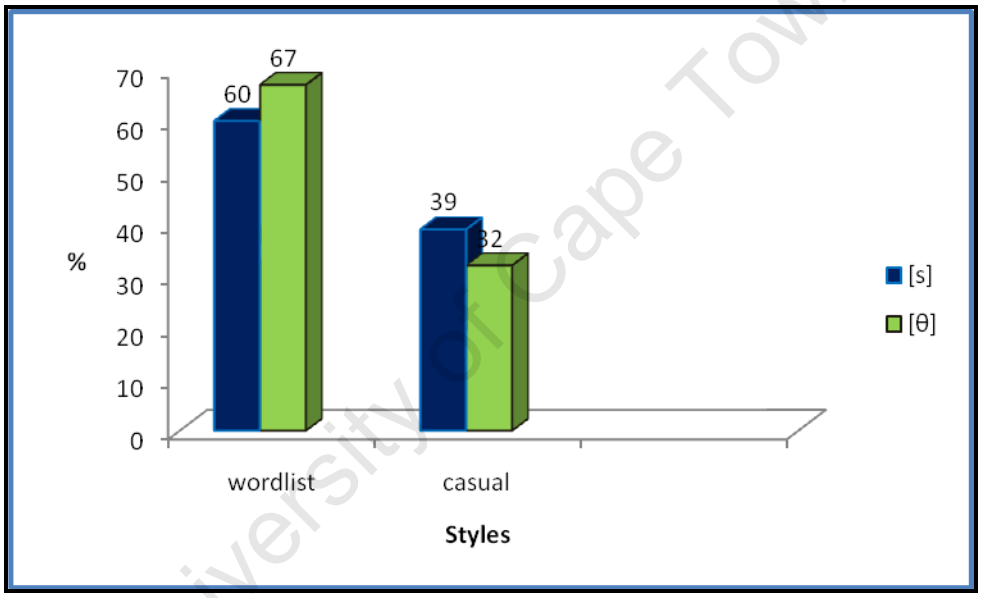
Both casual and wordlist styles had high percentages in the use of [z] against [ð]: 66% of 796 tokens in wordlist style and 61% of 316 tokens in casual style. Regardless of the occurrence of fewer tokens in casual conversation compared to wordlist style, the percentage use of variant [z] is high in both styles compared to the use of [ð], as graph 5.15 below shows.

Graph 5. 15: Percentage distribution of [z] and [ð] by style



Concerning [s] and [θ], 60% of 794 tokens in wordlist style and 67% of 901 tokens from casual speech favoured the use of [s] against [θ], as graph 5.16 illustrates. The percentage of [s] is higher for casual style and the percentage of [z] for wordlist style. Despite the difference in the number of raw tokens, the percentage difference in favour of [s] in both styles is minimal. This suggests that, in both styles, standard variants are used more frequently than non-standard variants. However, in the next stage the statistics will show further significant style.

Graph 5. 16: Percentage distribution of [s] and [θ] by style



5.3 Goldvarb analysis

Percentage results are not enough to describe how social factors influence the use of (z) and (s). Since - as mentioned earlier - some of the factors overlap, Goldvarb performs another phase of statistic analysis using a model that teases out which are the factor groups which best account for the data in relation to the dependent variables. Goldvarb calculates the values or factor weights for each factor group, showing how significantly each contributes to the model of variation, the loglikelihood, the significant

values and the maximum possible likelihood. It carries out a *binomial step-up/step-down analysis.* This involves running each factor group by adding each of the other groups to the analysis, one at a time, until all factors are included. It then performs a second level of analysis by removing one factor group at a time until only one factor remains. The best two stepping-ups and down models are normally identical, indicating the factor groups which have significant weight, with probable contexts regulating the occurrence of dependent variables. The eliminated factor groups do not significantly contribute to the variation. The strength of weight that influences the occurrence of the linguistic variation is measured by values between 0 and 1.00. A weight between 0.50 and 1.00 indicates that the factor favours the use of a variant relative to other factors in the same group. The value 0.50 is neutral and neither favours nor disfavours the dependent variable, while a value below 0.50 disfavours the variant (Bayley 2002:124-134, Paolillo 2002:34, Tagliamonte 2006:248). The model after the step-ups and downs is undoubtedly better than other models in the analysis if the 'significance' value is $p< 0.05$, but the question is whether it is the best model for the data. Through the 'Show model fit' function, the Goldvarb program can perform a G^2 test which compares the likelihood of the maximum possible model for the data with the model likelihood for every model. The results of this G^2 test is called 'Fit: X-square model-fit' in Goldvarb analyses. A model can only be considered a good fit for the data if this test returns an 'accepted' result, i.e. when the Fit: X square value has a p-value of $p>0.05$ (cf. Sigley 1997 & 2003, as cited by Hoffmann 2006:8). So a model can be selected as the best among other models, but at the same time can either be accepted or rejected as a best model fit for the data depending on this p-value. A model is rejected when there is a high interaction, which may be found across social variables, between linguistic and non-linguistic factors.

Based on these criteria, table 5.14 is the result for the best models and best model fit for the data, including factor weights from each factor group based on the best step-ups and step-downs in favour of [z] against [ð] and [s] over [θ]. The best stepping-up run for [s] is 28, and the best stepping-down run is 55. The best stepping-up run is 23 and the best stepping-down run is 57 for [z]. Table 5.14 also includes significance values for models, log likelihood, maximum possible likelihood, Fit: X-square, p-value for model

fit, and whether rejected or accepted. The results for all iterations are attached as appendices 5 and 6.

Table 5. 14: Internal and external significance factors in favour of [z] and [s]

Application value	**[z]**		**[s]**	
Log likelihood	**-626.128**		**-1020.525**	
Significance	**0.000**		**0.004**	
Maximum possible likelihood	**-610.240**		**954.549**	
Fit: X-square (29)	**31.778**		**149.952**	
p-value	**0.1272**		**0.0000**	
Model fit for data	**Accepted**		**Rejected**	
Total N	**1134**		**1695**	
Variable factors	**Factor weight**	**%**	**Factor weight**	**%**
Internal Factors				
The position of the syllable				
initial	0.75	89	0.66	78
medial	0.49	65	0.42	58
final	0.41	54	0.39	53
The following vowel				
back	0.63	78	0.46	60
front	0.36	51	0.54	67
Status of lexicon				
borrowed from Swahili	[ ]	64	0.57	72
native Chasu	[ ]	65	0.47	60
External factors				
Age				
young 21-40	0.61	75	[ ]	65
middle 41-60	0.45	56	[ ]	64
old 61+	0.41	58	[ ]	61
Sex/gender				
female	[ ]	64	0.44	59
male	[ ]	65	0.55	67
Education levels				
highly educated	0.67	79	0.61	69
less educated	0.41	57	0.44	58
Social class				
middle class	[ ]	74	[ ]	69
lower class	[ ]	54	[ ]	58
Styles				
wordlist style	[ ]	66	[ ]	60
casual style	[ ]	61	[ ]	67

The factor groups shown in table 5.14 above are those which were selected by the multivariate Goldvarb analysis as favouring the dependent variables [z] and [s]. The factors which favour [z] against [ð] are in column 2 of table 5.14. These include internal factors such as the syllable position, i.e. initial, middle, and final syllable, and the type of the vowels which follow the variants, i.e. front or back vowels. External factors were education levels and age groups. Factor groups eliminated (shown in empty bracket [ ]) during stepping up and down are 3, 4, 6, and 7, which represent the status of the lexical item (Chasu or borrowed from Swahili), gender, social class and style. This implies that these factor groups have an insignificant influence on the occurrence of [z]. The model was best among other models and was also 'accepted' as the best model fit for the data, as p-value was above 0.05.

The factor groups which were selected by the multivariate Goldvarb analysis as favouring the occurrence of variant [s] against [θ] are presented in column 4 of table 5.14. As with [z], factors such as the position of the syllable, and the vowel after the variants were selected. In addition, the status of the lexical item, i.e. whether from native Chasu or borrowed from Swahili, was also among the internal factors selected during the stepping up and down. The analysis also singled out external factors such as gender and education level. The multivariate analysis eliminated factor groups 5, 6 and 7, which are for age groups, social class (lower or middle) and style (also shown by empty bracket [ ]). While the factor age has an effect on variable [z] and not gender, for the variable [s] gender is a significant factor rather than age difference. However, this model was selected as the best model among other models, but rejected as a model fit for the data, as p-value is lower than 0.05.

5.4 Possible interactions

Another attempt was therefore made to find out to what extent there were interactions between independent variables, through a two-way factor group comparison, as in tables 5.15 and 5.16 below. Here, I was greatly assisted by Dr. Thomas Hoffmann, a visiting linguist from the University of Regensburg with special expertise in statistical analysis.

Table 5. 15: Two way factor group comparisons for (z)

Factors	Fit: X- square	Model fit status	P-value
Syllable & vowel	(3) = 7.044	accepted	0.0746
Syllable & lexicon	(3) = 6.557	accepted	0.0902
Syllable & gender	(3) = 1.137	accepted	0.1508
Syllable & social class	(3) = 0.628	accepted	0.0767
Vowel & education	(2) = 0.039	accepted	0.9808
Age & style	(3) = 9.105	rejected	0.0309
Lexicon & education	(2) = 0.003	accepted	0.9987
Gender & education	(2) = 2.449	accepted	0.2950
Style & education	(2) = 0.492	accepted	0.7829
Social class & style	(2) = 4.193	accepted	0.1298
Lexicon & social class	(2) = 0.455	accepted	0.7965
Gender & social class	(2) = 0.345	accepted	0.8429
Vowel & social class	(2) = 0.086	accepted	0.9584
Gender & vowel	(2) = 4.302	accepted	0.1218
Lexicon & gender	(2) = 3.702	accepted	0.1652
Education & age	(3) = 5.060	accepted	0.1740
Age & gender	(3) = 1.874	accepted	0.6044
Vowel & age	(3) = 2.358	accepted	0.5017
Syllable & age	(5) = 4.128	accepted	0.5330
Syllable & style	(3) = 9.431	rejected	0.0260
Lexicon & age	(3) = 2.862	accepted	0.4236
Lexicon & style	(1) = 0.02	accepted	0.9701
Vowel & style	(2) = 17.386	rejected	0.0002
Social class & education	(1) = 0.004	accepted	0.9484
Gender & style	(2) = 1.3228	accepted	0.5172
Vowel & lexicon	(2) = 1.979	accepted	0.3839
Age & social class	(3) = 8.868	rejected	0.0344

From table 5.15 we can see that models are accepted except where style difference is compared to social class, syllable position, following vowel and age difference. As mentioned earlier, the model is rejected where factor groups interact highly with each other.

Table 5. 16: Two-way factor group comparisons for (s)

Factors	Fit: X- square	Model fit status	P-value
Syllable & vowels	(3) = 28.910	Rejected	0.000
Syllable & lexicon	(3) = 38.15	Rejected	0.0000
Syllable & age	(5) = 8.315	Accepted	0.1473
Syllable & education	(3) = 2.869	Accepted	0.4225
Vowel &education	(2) = 1.206	Accepted	0.5535
Age & education	(3) = 18.943	Rejected	0.0003
Vowel & social class	(2) = 4.12	Accepted	0.4950
Vowel & style	(2) = 15.619	Rejected	0.0004
Vowel & age	(3) = 3.715	Accepted	0.2949
Vowel & gender	(2) = 2.026	Accepted	0.3747
Vowel & lexicon	(3) = 0.616	Accepted	0.7365
Syllable & style	(3) = 68.482	Rejected	0.0000
Age & social class	(3) = 11.619	Rejected	0.0095
Lexicon & education	(2) = 1.016	Accepted	0.6099
Gender & education	(2) = 1.623	Accepted	0.4537
Education & social class	(3) = 0.009	Accepted	0.9282
Syllable & gender	(3) = 5.434	Accepted	0.1508
Gender & lexicon	(2) = 9.771	Rejected	0.0089
Age & gender	(3) = 54.940	Rejected	0.0000
Lexicon & style	(1) = 0.006	Accepted	0.9403
Gender & style	(2) = 9.734	Rejected	0.0090
Age & style	(3) = 4.421	Accepted	0.2227
Lexicon & social class	(2) = 1.046	Accepted	0.6009
Gender & social class	(2) = 3.687	Accepted	0.1662
Social class & style	(2) = 0.543	Accepted	0.7637
Style & education	(2) = 0.18	Accepted	0.9103

In table 5.16, we can see that models are accepted, except those for gender and style, gender and age difference, syllable position and status of lexical item, and age and

educational levels. This implies there is high interaction between those factors with rejected models.

A further attempt to find out the accepted model fit for the data was done with the assistance of Dr. Thomas Hoffmann from the University of Regensburg, who was a guest lecturer at the University of Cape Town. Dr. Hoffmann introduced me to the statistical program R. R is a language and environment to implement statistical computing and graphic techniques. A scholar, Johnson[3] , designed a manual for Rbrul which he describes as a program used by linguists to analyse sociolinguistic data in the R environment. According to him, Rbrul works better and faster, and resolves analytic problems where Goldvarb falls short. One of the Goldvarb shortcomings is its inability to handle efficiently interactions between factors, especially where they are assumed to be independent of each other. Such interactions (and empty cells) were unfortunately found in my data, on account of the skewed nature of occupations by age, social class and education, as discussed earlier. Moreover, there were interactions between age, social class and education themselves. Johnson (2009:5) argues that Goldvarb tends to ignore the grouping and treats each token as if it were an independent observation while in reality linguistics data sets are 'naturally grouped according to the individual speakers who produced them'. Through Rbrul analysis, model 19 (cf. appendix 6) appeared at least to be the best Goldvarb model fit for the data and was 'accepted'. In this model, only three factors were selected as significant enough to influence the use of [s] against [θ]. These are syllable position, gender difference and educational attainment, as table 5.17 illustrates.

[3] *www.ling.upenn.edu/~johnson4/**Rbrul**_manual.html*

Table 5. 17: Internal and external significant factors in favour of [s]

Application value	**[s]**	
Log likelihood	**-1029, 87**	
Significance	**0,000**	
Maximum possible likelihood	**-1023, 211**	
Fit: X-square (8)	**13, 322**	
p-value	**0,1017**	
Model fit for data	**Accepted**	
Total N	**1695**	
Variable factors	**Factor weight**	**%**
Internal Factors		
The position of the syllable		
initial	0.66	78
medial	0.42	58
final	0.37	53
The following vowel		
back	[ ]	60
front	[ ]	67
Status of lexicon		
borrowed from Swahili	[ ]	72
native Chasu	[ ]	60
External factors		
Age		
young 21-40	[ ]	65
middle 41-60	[ ]	64
old 61+	[ ]	61
Sex/gender		
female	0.44	59
male	0.55	67
Education levels		
highly educated	0.61	69
less educated	0.43	58
Social class		
middle class	[ ]	69
lower class	[ ]	58
Styles		
wordlist style	[ ]	60
casual style	[ ]	67

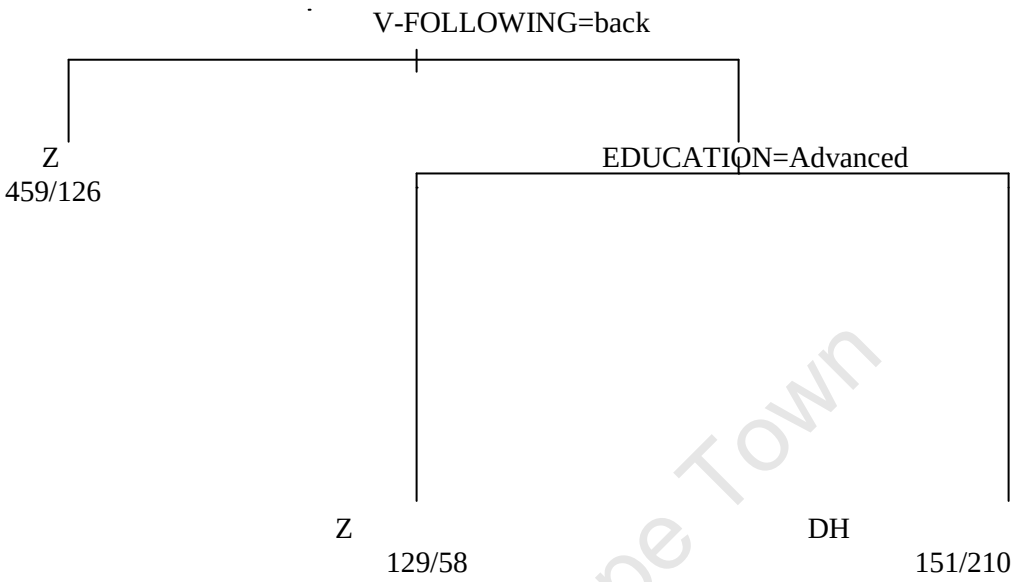
While the Fit: X square test only shows that this model was a good fit for the data, the Nagelkerke R^2 value shows how well the model fits the data, i.e. how much of the data is actually explained by a model. In this case, the three variables in table 5.17 together only account for 0.117, which is about 12%. This means that the model explains only 12% of the data, which is not significantly impressive, and the main attributes could partly be the interactions within factors, and the fact that, generally [s] appears to be strongly favoured over [θ].

5.5 CART Analysis

Finally, a CART Analysis (Classification and Regression tree) was used to describe the data. According to Baayen (2008:148-154), CART trees in R environment involve binary splits on variables to establish whether the class of data set can be predicted. It involves the formation of leaves from each node, and each leaf contains the best prediction based on the set of data. Splits in the CART tree are labelled with decision rule and are recursively formed. The following CART tree models, in figures 5.1 and 5.2, were made with the assistance of Dr. Thomas Hoffmann.

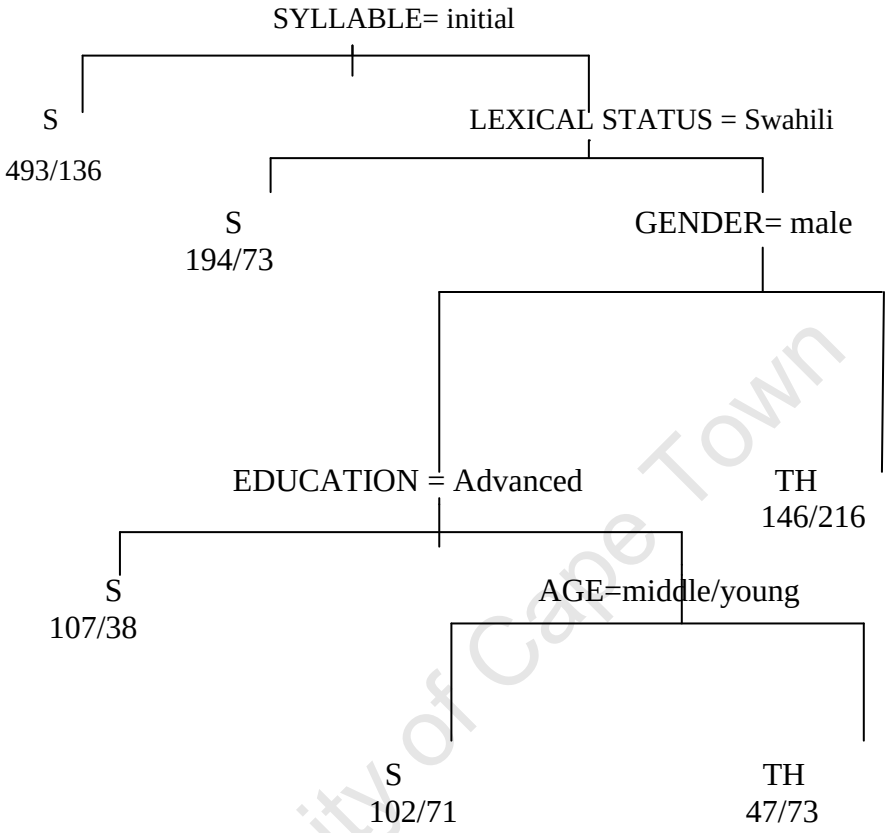
The CART tree model in figure 5.1 is describing [z] against [ð], while figure 5.2 describes [s] over [θ].

Figure 5. 1: CART tree model for [z] against [ð]



The overall data distribution for [z] is 65.2% and 34.8% for [ð]. While the null model for [z] against [ð] thus predicts 65.2% correctly, the CART model predicts 70.4% correctly. Two factors were used to build up splits, which help to predict model fit: the vowel following the variants, and educational attainment.

Figure 5. 2: CART-tree model for [s] against [θ]



With variable (s), the overall distribution of [s] over [θ] is 64% and 34% respectively. While the null model predicts 64% correctly, the CART model predicts 70% correctly. The CART model for (s) involves five independent factors, and the way they interact with each other. These are syllable position, lexical status, gender, educational attainment and age difference. According to Baayen (2008:148-150), the information in the splits of the CART-tree is accessed by following the left branch first, and then the right until all the sub-trees are reached. Each branch/leaf node specifies a partition of the data set, which, together with other branches, forms a complete set of data in the root node.

5.5.1 Vowel following the variable

Concerning the environment of the vowel following (z), figure 5.1 indicates that [z] is supported by 459 observations of back vowels /a/ and /o/; and contradicted by 126 observations of front vowels /i/ and /e/. This implies that (z) is more likely to occur as variant [z] when it is followed by back vowels, regardless of any other variable, and less likely to occur with front vowels; while variant [ð] is favoured only when the variable is followed by front vowels and uttered by less educated speakers. On the other hand, with variable (s) this factor does not have a significant effect on the occurrence of [s] against [θ] as far as figure 5.2 is concerned.

5.5.2 Syllable position

Concerning the environment of the syllable, figure 5.2 illustrates that variant [s] occurs as an initial syllable in 493 observations, while 135 observations contradict. Normally, CART analysis splits into binary partitions and selects only those sets of data which favour most the dependent variable. In this case, initial position was compared to middle and final positions, which appear to favour [θ] most. This implies that in Southern Chasu word formation, (s) is likely to occur as a voiceless alveolar fricative [s] when it occupies the initial syllable position, while variant [θ] most occurs in the middle and word final syllable positions. The factor relating to the environment of the syllable, on the other hand, does not appear to affect the occurrence of [z] over [ð].

5.5.3 Status of lexical item

Turning to status of the lexical item, figure 5.3 shows that [s] occurs in borrowed words from Swahili in 194 observations, while only 73 contradict, meaning that [s] occurs most in borrowed words from Swahili. Conversely, variant [θ] is favoured in native Chasu words rather than in borrowed words from Swahili. This also supports the hypothesis that, although there are a few borrowed words from Arabic via Swahili which contain variant [θ] (as mentioned earlier as well as in chapter 4), the occurrence of [θ] in

Chasu is likely to be generated from within itself own or from neighbouring Northern varieties rather than Swahili.

Unlike for variable (s), this factor of the borrowed or native status of lexical items is an insignificant factor for (z). Previously I have shown that, of a total 194 tokens from Swahili, 63% favoured [z]. Out of 918 tokens from Chasu, 66% favour [z]. This means that [z] occurs in Chasu native words in almost the same frequency with Swahili borrowed words- i.e. the difference is insignificant.

5.5.4 Gender

Concerning gender, figure 5.2 illustrates its significant effect on the use of variable (s). The CART-tree shows that if the syllable is in the middle or final position in a native Chasu word, women's use of [θ] is favoured by 146 observations only, while 216 observations contradict. Age difference and educational levels do not affect women. On the other hand, if the syllable is initial position and the lexical item is borrowed from Swahili, men are likely to use variant [s] subject to their age and educational level. This means that women's use of [θ] is higher compared to men who use [s].

With regards to variable (z), gender difference is an insignificant factor in favour of [z]. As shown previously, the percentage distribution among males and females in favour of [z] is 66% from 398 tokens and 64% from 332 tokens respectively. This means that both men and women speakers are using [z] against [ð] in almost equal frequency. As a result, the factor was eliminated.

5.5.5 Age groups

Pertaining to the age group, the CART-tree analysis in figure 5.2 indicates that young speakers from 21-40, and middle-aged speakers from 41-60 favour the use of variant [s] by 102 observations with 71 contradictions. This pattern occurs only if the syllable is in the initial position in a lexical item borrowed from Swahili and by a young

or middle-aged, highly educated speaker. On the other hand, if the syllable is middle or final position in a native Chasu word, less educated old men's use of [θ] is only 47 observations against 73 contradictions. Unlike in the analysis of variable (s), where younger speakers are in favour of variant [s], age was an insignificant factor for variable (z).

5.5.6 Educational levels

On the other hand, with regards to educational levels, figure 5.1 indicates by 129 observations contradicted by only 58 observations that highy educated speakers favour the use of variant [z]. On the other hand, the variant [ð] is favoured by less educated speakers only if uttered with back vowels, indicated by 151 observations with 210 contradictions. This means that the latter favour the use of [ð] over [z].

Likewise, with variable (s), figure 5.2 illustrates how gender difference interacts with education in affecting the use of [s] and [θ]. Men's use of [s] is highly connected with advanced educated speakers' use of [s], by 107 observations against only 38 contradictions. This signifies that highly educated men use [s] more, while less educated women favour [θ]. Generally, the CART-tree analysis for (z) and (s) shows that the more educated the speaker is, the greater his/her chances of using the prestigious variants [s] and [z]. However, education matters only subject to different structural constraints.

5.5.7 Social class and style

While in the urban Western societies, social class differentiation and style offer a routine way of accounting for linguistic variation, the Goldvarb as well as Rbrul through CART-tree analyses of data in this study have shown that social class and style confer less effect in the linguistic variation of variables (s) and (z).

5.6 Discussion

5.6.1 Interaction between age groups and education

Sociolinguistic studies stress the important role that age plays in language change (Trudgill 1986, Chambers 1992, Romaine 1995). Although adults do have a capability for innovation, language innovation tends to be widespread among the youth (Eckert 1997:163, 2000:16). In this study, the Varbrul analysis does not give age as a significant factor for (s) [cf. Tables 5.14]. The CART-tree for (z) shows that age does not play a significant role [cf. Figure 5.1]. The CART analysis for (s) [cf. 5.2] shows that age is only relevant as part of a complex interaction. Age plays a role only in native Chasu words if the syllable position is non-initial; and the sub-group of older, less educated men favour [θ] while the young and middle-aged speakers favour [s]. The CART-tree in figure 5.2 indicates that, when (s) occurs, an initial syllable position is mostly realised as variant [s], especially in words which are borrowed from Swahili rather than native to Chasu. It also indicates that speakers who mostly use [s] are men from the young and middle age groups who have attained a high level of education. The reverse is also true: women from older groups who are less educated mostly favour [θ]. This signifies that there is an interaction between age difference and educational level in using standard variants. Here the historical background of the education system and the language of instruction in Tanzania are relevant factors.

Like any other Tanzanian rural society, the Asu people have passed through various phases of Tanzania's education system, during colonialism, during the struggle for independence and after independence. This periodisation is reflected in the linguistic results of the age groups of our study.

***The colonial period*:** During colonialism, the Tanzanian education system had several levels: (a) starting with grade 1 to 4 at the lower primary level, (b) after passing a national examination, one could go to middle school up to grade 8, and (c) lower secondary school from grade 9 to 12, and 13 to 14 for advanced secondary school. Education from middle

school onward was subject to passing the examination and the capacity to pay school fees. Lack of finances led to very few children continuing with middle school, let alone secondary schooling.

Alidou (2004) states that, before the arrival of colonialism, choosing the language of instruction was not a problem, since education was community based and orally transmitted: each community could use its own language to educate its children in its own way. It was during colonialism that education across ethno-linguistic groups was introduced. The problem of the language of instruction started as children of different language backgrounds were placed in one class in formal schools for Western education. The language of instruction was also determined by the objectives set in providing education for the Tanzanians. During the German colonial period, the goal of education in government schools was to prepare Tanzanians for employment in the colonial bureaucracy. Hence Swahili was taken to be a convenient and practical lingua franca, as it was spoken by nearly all potential employees in schools and in colonial administration (Roy-Campbell 2001:41-42). While government schools insisted on using Swahili as a medium of instruction, mission schools which exceeded the government schools both in number and influence used Swahili with the local language of a particular ethnic community.

In Vuasu, particularly Same District, community members especially the elders, claim that mission schools during the German colonial rule used Swahili and Chasu simultaneously. Early mission schools were established by German missionaries, through the Seventh Day Adventist Church, which was introduced for the first time in 1903 in Mamba Giti. The latter was then called Friednstal Mission station, the first Adventist station established by Germans in Tanzania. This mission was followed by another one in Kihurio in 1905 (Hoschele 2005) and then Suji in 1907. The motivation behind colonial education in these mission schools, particularly in Vuasu areas, was to transmit skills in carpentry, tailoring and needlework, agriculture and shoe-making. This needed new converts and people who could learn languages and serve as evangelists, teachers, "house boys" and translation assistants. Hoschele (2005) narrates the biography of one of the

new converts in the early church, Paulo Kiame Risase. He worked as a secretary for the mission for a short time, and then became a teacher in 1910, teaching at the schools in Giti and Vunta, both located in his native South Pare Mountains. In 1914 he worked as a full-time assistant to Ernst Kotz, translating the New Testament into Pare (Asu) and personally typing the manuscript.

Mission schools like Giti, Suji, Kihurio, Vunta, Bwambo and Vugwama, which were owned by the Seventh Day Adventists, and Kiranga and Gonja owned by the Lutheran Church, used both Chasu and Swahili as the media of instruction. Giti and Kiranga were located in one of the research areas, Mamba Myamba. Rubagumya (1990) states that, during British colonial rule, Swahili was preserved as a medium of instruction in government schools in the first five years of primary school. In the last three years of primary school and all of secondary school, English was used, as the British colonial administration was carried out in English. The British colonial government, as Roy-Campbell (2001) argues, had a concrete plan to train a minimal number of Tanzanian elites who could assist in colonial administration, and to maintain very low levels of education for the rest of the population.

In the Vuasu area, where mission schools exceeded government schools in number and influence, Swahili and Chasu were the media of instruction even during the British colonial rule for lower primary education. Those who could afford the fees went to Parane middle school, where English was introduced and Chasu was no longer used. So very few of the older speakers who had attended schools, especially during colonialism, were exposed to Southern Chasu as their medium of instruction in schools. In the church, they used Southern Chasu dialect Bibles and hymn books, which included the variants [s] and [z] instead of [ð] and [θ]. However, the overwhelmingly illiterate majority in the society retained both variants, especially in informal speech. This is because very few managed to attend school up to grade four. We have explained above that groups of Vaasu people came from the southern areas with variables (s) and (z), which are evident in the southern neighbouring languages. Still other Vaasu moved from the northern parts with [ð], which is still evident in the northern neighbouring languages.

Though it is difficult to ascertain which variant is older than the other, it is worth acknowledging that, for a certain period of time, these two variants were used simultaneously. Periodically, those who attended school were in a position to use standard variants [s] and [z] more frequently as they were used in school. However, in the Chasu society very few old speakers are left, and their contribution to language change is negligible.

During the struggle for and in the years immediately after independence, as explained in chapter one, more emphasis was given to Swahili. Swahili was used all over the country for political campaigns during the struggle for independence. After independence, all mission schools were nationalized and a few abandoned, as the newly independent Tanganyika tried to find new directions economically and politically. Schools in rural areas declined, as the government could not run all the schools which were previously owned by missionaries in remote rural areas. The level and quality of education started to deteriorate at the national level, especially in the rural areas, which received less financial assistance than urban centres. In explaining the different phases of education and economic development in Tanzania after independence, Galabawa (2001:8) argues that the country experienced a long and deep economic crisis during the 1970s and 1980s. The unfavourable economic trends reduced the capacity of the government to finance social services, including education. From 1967 to the 1980s, Tanzania used the ideology of Socialism and Self Reliance (SSR). Though the Universal Primary Education (UPE) policy was introduced during this period, it was inefficiently sustained. For a poor country like Tanzania, which was committed to a call to equitable and fair provision of education for all, UPE could be a reasonable expectation. However, Galabawa reports that UPE received very little support from international donor agencies, and hence had to be solely supported by the Tanzanian government. Though there is evidence of high rates of enrolment and intake into primary education during the implementation of UPE, the system was insufficiently attended internally and proved inefficient. This is the time when the middle-aged speakers of our study were raised and went to schools that had high rates of enrolment, but with inadequate services. Others remained at home doing cultivation, which was highly emphasised at that time by the

policy of SSR. One could perceive this time as a confusing period, especially for the rural people as far as language use was concerned. Swahili was emphasized for nation building and to eradicate ethnicity, and ethnic languages were no longer certified for use in lower primary schools. Some changes that can now be seen in the Tanzania ECLs were initiated during this period. Chasu was subsequently only used in conversations, rather than formally in schools as a medium of instruction. Though Swahili replaced Chasu, it could be accessed only by those who went to school. So Chasu was regularly used by a large number of people, though it was not the standard variety previously taught in school. Swahili Bibles and hymn books started to be used in the church in place of Chasu. However, the highly educated middle-aged group maintains [s] not because they used Chasu as their medium of instruction; rather, they are aware of the status of this standard variant [s] and the stereotype associated with less prestigious variant [θ], and they have exposure to the formal settings of language use like in public meetings, church areas.

Galabawa describes the 1980s-1990s as the period of growth of new economic collective national thought. The period is characterised by the introduction of liberal ideas of free market choice and cost efficiency related to schooling. During this phase, together with other national priorities, UPE was expanded. However, since the national economy was not stable enough to sustain all of these provisions, this phase was also negatively affected by the Structural Adjustment policies which surfaced in the mid-1980s. This period was subsequently characterised by falling enrolment rates and high internal inefficiency. Another phase, which commenced in 1995, is marked by what Galabawa (2001:8) terms 'income and non-income poverty collective thought', characterised by (i) 'a donor dependency syndrome and defeatist developmental mindset, (ii) a weak and low capacity for education management, (iii) failure in good governance in the organisation and production of UPE and (iv) ineffective policy implementation'. In these later periods of time, education was perceived as an asset to compete for. The affluent people could compete for the best-rated education at high-quality schools, while the poorer remained in poorly resourced schools. Swahili and English have been used throughout this period, while language is also considered a ladder to higher economic standing. A high status language is associated with a good standard of living. If one cannot compete for a better

school or higher education, one can at least try to master a highly esteemed language. This is the time when the young speakers of our study were raised and went to schools. We have seen in chapter 4 that, in rural settings, Swahili is not only taken as a national language, but is additionally associated with modernity and urban life. It is seen as a variety for young people in opposition to Chasu, which is associated with older speakers, rural settings and an imagined traditional world. Since Swahili is written with the same script as Chasu, there is the potential for mutual influence in reading style. Swahili orthographic <z> and <s> represent [z] and [s] respectively. Those Chasu speakers who are literate in Swahili then transfer this value to [z] and [s] as the standard variants of variables (z) and (s) in the Southern Chasu dialect, instead of stereotyped variants [ð] and [θ]. This orthographic influence appears to have been carried into both formal and informal speech as well. However, observation shows that standard and non-standard variants are used interchangeably, especially in informal settings.

Though social class does not seem to affect phonological variation directly in this chapter, it was explained in chapters one, three and four that the distribution of the economic activities in this rural setting correlates with different age groups. It is also connected to the norms of those who traditionally own the land or animals as their properties. It is also related to the unique ways that economic and social development have impacted upon young people's attitudes and expectations. We have seen that this society is patriarchal; therefore the father, as the head of the household, owns the properties. The sons will inherit the land or animals upon the death of their father, and the cycle continues throughout generations. Occasionally, the land for inheritance becomes too small for further division, necessitating that individuals engage in other activities, with youngsters setting off for business or professional activities. In this society, business is supposedly for the young and a few middle-age speakers, as they are felt to be sufficiently strong to travel frequently and build external networks. Swahili is often the medium of communication in such networks. It was discovered in chapter four that young speakers borrow more frequently from Swahili and code-switch as a result of their interactions in business and professional activities, which are the contemporary options for youngsters. Those who engage in peasant activities cultivate ginger and sugarcane as

cash crops in this rural area for the time being. After harvesting, they sell or transport their cash crops to the city, mostly in Dar-es-salaam, Arusha and Mwanza. Unlike the older generations, young people face the prospect of being forced out of the village to the town, due to economic changes. Due to increased mobility into and out of the village, there has been an increasing sense of loss of traditional practices attributed to the arrival of the outsiders. As a result, town life has directly impacted upon the young Vaasu, and their speech forms are being affected. It is unsurprising that some of the Swahili elements have filtered into young people's speech. The CART-tree in figure 5.2 has shown that [s] is likely to occur more frequently in borrowed words from Swahili compared to lexemes from Chasu where [s] and [θ] co-occur. This is a result of language contact, which comes with the level of education and the activities in which young and middle-age speakers tend to engage.

In this study, very few old educated speakers who maintain [s] conserve the features of dialect simultaneously (they hardly code-switched, as discussed in chapter 4), though they use the variants interchangeably. The middle-aged group is characterised by a transition period, from a decline in using both dialect and the standard variants (old-aged educated speakers) to the rise of using standard variants but jettisoning the dialect variant itself, due to the influence of Swahili among young speakers. Different from other age groups, youngsters preserve [s] of Southern Chasu because they are sensitive to the stigmatized features like [θ] of their parent's variety. They do not, however, conserve many forms from Southern Chasu, because they use Swahili during their working hours and Chasu at home with the family. So one can generalize that young and middle-aged speakers use [s] rather than [θ] due to their level of education, and also in this context as a result of their exposure to Swahili, and the structural constraints of Chasu words. There is thus a change in progress through young and middle-age speakers in reducing the use of the stigmatized feature [θ] in Southern Chasu, as a result of using Swahili and only with the effect of syllable position. This change in norms of middle-age and younger speakers has a different meaning compared to the study of diphthong centralization among the residents of Martha's Vineyard. Labov comments that increased centralization of (ay) and (aw), which is the feature of Vineyard speech, is interpreted as an expression of strong

resistance to the "invasion" of summer people; i.e. wealthier people from the mainland who are ever-present in summer residences. Labov (1972:28) adds that "when a man (Vineyarder's) says [raɪt] or [haʊs], he is unconsciously establishing the fact that he belongs to the island: that he is one of the natives to whom the island really belongs". Contrarily, young and middle –aged, highly educated rural speakers from Southern Chasu use the standard variant to express their willingness to embrace and comply with the town-oriented way of life and new culture through Swahili.

5.6.2 Gender

Modern gender studies as we have seen in chapter two have often shown that, in the industrialized societies where social stratification is revealed through speech, women tend to use standard variants significantly more frequently than men, who are prone to stigmatised features (Wolfram 1969:76, Labov 1972a:243, Trudgill 1974a:93, Cameroon & Coates 1988:13, Chambers 1992, Cheshire et al 1999). The tendency of women to prefer more of the standard form is reflected in "their conservatism, prestige consciousness, upward mobility, insecurity, deference, nurture, emotional expressivity, connectedness, sensitivity to others and solidarity. Men's language is heard as evincing their toughness, lack of effect, competitiveness, independence, hierarchy and control" (Eckert and McConnell-Ginet 1992:90), as these perceptions are positively valued in certain male circles. Labov (1998:19) points out that the mechanism of change is not connected to sex difference in any clear and simple way, as either sex can be the dominant factor for using a certain variable.

The data has shown that men use the standard variant [s] of Southern Chasu more frequently than women. This is contrary to the expectation that it is women who use standard variants more frequently in Western societies. As explained in chapter one, this rural community is patriarchal, whereby men embrace greater power than women. My personal observation is that, secularly, men hold power at all levels: from the grass roots, to the household and *kitongoji* (an area with ten households), to the village, ward and division level. This practice is inherited from the formerly traditional leadership of

Mfumwa, the chief who was most commonly a man. Few women attend political or public meetings, or any gathering that discusses rural development. Women who go to these gatherings seldom speak, because they understand and comply with the custom that talking in public is not meant for women. It is the father who speaks with the teacher regarding the progress of the children in school. He is a representative of the household in all socio-economic matters.

Regarding religion, it was traditionally men who could lead the ritual procession in the shrine *itasio.* This has not changed with the advent of Western religions or Islam, which emphasise men as leaders. In church, women do not participate in making decisions, as they don't speak much. It is through these gatherings that men practise and display mastery of the standard form of the language or variety. Women are denied this opportunity to perfect their use of the standard variety. They are denied networking opportunities with people who practise using language in public formal settings.

Socially, women who dare to speak in a gathering are stereotyped not just by men but by their fellow women. They are considered to be not well groomed by their mothers. Men normally scorn such women with this saying: "*Siovoa mche ena kiteto"*, 'I don't marry a talkative woman' (who dares to speak in the public gatherings). As a result, women use both variants [s] and [θ] in most of their informal conversations, while some of them can hardly differentiate the standard vs. non-standard variants. As a result, this tendency reverses the Principle I we discussed in chapter two: "In a stable sociolinguistic stratification, men use a high frequency of non-standard forms than women" (Labov 1998:7). However, this tendency is not the case with variable (z), where both male and female use standard variant [z] almost in the same frequency. This implies that the occurrence of (s) and (z) are as a result of different influences within the same community. Though social factors may contribute, the structural constraints are much more powerful, as CART-trees in figures 5.1 and 5.2 illustrate. The environment of vowels influences the occurrence of variant [z] over [ð], while syllable position and the status of the lexical item influence the occurrence of either [s] or [θ]. Meanwhile, changes with variables (z) and (s) do not occur simulitaneously, implying a kind of transition in

the society as far as language variation is concerned, or that the data for variable (z) is not large enough to predict gender on its own as a factor underlying use of one variant over the other.

Women in Western societies are more conscious of and insecure as regards their social status, and consequently use linguistic features to reflect their social status. They use standard forms as an expression of "symbolic capital" (Eckert 1989:256) or rather as constituting a strategy of "self promotion" (Gordon 1997:48), since standard features are associated with people who possess socio-economic power in the society. Observation shows that women in the Chasu rural setting still believe in the traditional system where it is the responsibility of their husbands to speak for the household. Husbands are justified in prospering and acquiring higher economic status for the betterment of the family. Women consider it worthwhile to struggle to marry a wealthier man, and subsequently not to compete with his high socioeconomic status, especially through language, in which they run the risk of being measured as a mannerless woman.

5.6.3. Gender with education levels

The CART-trees in figures 5.1 and 5.2 have shown that the higher the education level the greater the tendency to use the prestigious variants [z] and [s]. Figure 5.2 has advanced that the tendency of men to favour the standard variant more than women can also be associated with educational levels and age difference. Women generally favour [θ] only if the syllable is non-initial and in Chasu words, while for men education and gender matter. It is highly educated, young and middle-age men who are likely to use variant [s] more frequently than women. [z] is favoured by highly educated speakers only if the following vowel is back, and [ð] is only favoured by less educated speakers if the following vowel is front and gender is not involved. Through education, one is exposed to standard language, which may result in a gradual decrease in the use of stereotyped linguistic features, regardless of sex difference. Access to standard features is necessary for acquisition, when the standard is different from daily conversational usage.

In the rural Vuasu setting, it is apparent that very few women go to school beyond the primary level, and as a result most of them are less educated. It is perceived that they are not responsible for the economy of the household. They don't possess or inherit land; rather, they are mostly peasants, working on their husband's farms. They are expected to be married after primary school rather than have a career outside marriage, and so are considered not to need simple skills beyond reading, writing and doing uncomplicated arithmetic, which they can acquire through primary education. During the colonial era, women who happened to go to school ended up in grade three or four. In the colonial era they finished in grade seven, as parents would prefer not to educate a girl who is to become a wife to someone beyond this level. Girls would thus marry in their mid- and late teens without attending secondary schools and colleges to any significant level. This could be the reason why women are lagging behind in using variant [s], as they were not exposed to school and thus could not learn intensively how to distinguish the standard from the non-standard variant. This tendency of men to favour standard features in Chasu society can be equated to studies in Muslim societies ((Modaressi 1978 & Abd-el-Jawad 1981) as cited by Labov (1998:12-14) and Chambers (2003:156-158)). In these societies, women play a minor role in public life, and are hence afforded less access to features of standard and or Classic Arabic. As a result, men make more use of standard speech, as they are able to access and apply it in public settings.

Labov (1998:13) mentions two other studies-by Hibiya (1988), who studied several variables in Tokyo, and Morale (1986), who studied the velarization of /n/ in Puerto Rican Spanish - in which no significant sex difference was found. Variable (z) in this study does not show any correlation with gender, in the sense that both men and women frequently make use of the standard variant [z]. Men scored a raw percentage value of 66%, while women scored 65% in using [z] against [ð].

It was discovered in chapter four that women code-switch more frequently than men, though the difference is not significant (t-tests between men and women are 0.515858 for Chasu-Swahili, 0.168968 for Chasu-English). This could be a result of transformation in the education process. In Tanzania, observation shows that since the

year 2000, the intake of females in secondary schools is rising, though numerically they are still fewer than men. Additional government secondary schools have been established in the rural areas to enable enrolment for both boys and girls. From the analysis of this data, we can speculate that, in future, all factors being equal, especially accessibility to education, women may use standard variants with greater frequency.

We have seen in chapter two that in Principles Ia and II women favour incoming prestige forms and at the same are innovators for certain changes from below. Women lead in the acquisition of the new prestigious features and the elimination of stigmatised patterns (Labov 1998:14). The importation of new prestige forms involves drawing in norms external to the speech community. In their network studies, Milroy & Milroy (1985) have shown that men and women tend to socialize differently, and this may be reflected in language variation. They also note that language change is arguably more probable through wide network contacts, for it is the speakers with more contact with people from outside the community who are likely to absorb external linguistic features. Language changes from both below and above in Chasu community involve men rather than women, since men have more external contact than women. We have seen in the introductory chapter that in this society, women are peasant housewives, while men are attached to professional jobs, businesses which involve travelling far and spending some days or weeks away from home. Women spend much of their time farming and raising children. They only engage in small business, mostly for vegetables and fruits. They hardly import any new features into Chasu, as they don't travel far or spend time with outside communities. However, more changes are anticipated in the future.

5.6.4 Gender and social class

The review in chapter two showed that social stratification in Western urban communities has been central to an analysis of sociolinguistic variables. It was established that, in many speech communities, especially where linguistic variables are stable, in the interaction between gender and social class, Lower Middle Class (LMC) women use a high proportion of prestige forms, are sensitive to new prestige variants and

are prone to hypercorrection (Coates 1986:64-72). LMC women copy features of the Middle Class (MC), whose language behaviour is more standard, in order to gain social prestige. However, through Rbrul and CART analysis in the sociolinguistic study of the Chasu speech community, class has been shown to be insignificant as a social factor for both variables (z) and (s).

Observation shows that, unlike the situation in most Western communities, a woman in rural society is still valued through symbolic capital associated with the community's norms and stereotypes for appropriate female behaviour like being married, being a good mother, a good wife, generous, a good cook, a hard worker (on the farm), modest and polite. However due to current economic changes, women tend to engage in small businesses and a few receive further education, which progressively changes their status in terms of occupation type.

5.6.5 Gender and age groups

Concerning gender/sex and age differences, the psychological literature gives evidence of women's verbal superiority (Chambers 2003:148). Over many years Western women have demonstrated an advantage over Western men in fluency, speaking, sentence complexity, analogy, and listening comprehension of written and of spoken materials, vocabulary and spelling (Maccoby and Jacklin 1974:75-85), and female precocity in verbal skills begins to show during their infancy. Research findings have demonstrated that, in Western communities where sex difference is established, women are found to be ahead in using standard variants since their childhood. This claim is related to neuropsychologists' hypothesis of sex-based brain asymmetry, which shows that women's brains appear to be more globally organised for specific functions, different from men's brains, which appear to be more highly lateralized, with verbal function in the left and spatial function in the right hemisphere (Kimura 1987, as cited by Chambers 2003:151). It is from this hypothesis that women are considered to be more advantaged in verbal skills and sociolinguistic competence. However, it should be noted that, in linguistic behaviours, speakers are influenced by their peers rather than their parents;

change occurs when speakers imitate the speech of sub-groups with whom they wish to identify (Coates 1986:149). This is evident in the Chasu community, as CART-tree figure 5.2 shows that it is young and middle-aged men who use the standard variants more than other groups; however, subject to structural constraints. This finding contrasts against gender studies in Western urban communities.

5.7 Chapter conclusion

This chapter has presented a quantitative analysis of variation between Southern Chasu dialectal forms of variables (z) and (s), in correlation with sociolinguistic variables such as gender, social class, style, age and education as external factors, and syllabic position, vowels after the variable sound, and the nature of the lexicon as internal factors. The statistical analysis to detect a model fit for the data was performed through Goldvarb analysis, and later with Rbrul, where Goldvarb proved to be unsuccessful because of interactions between different groups resulting in some empty cells.

It was discovered in this chapter that (z) and (s) occur respectively as either variant [z] or [ð] and as [s] or [θ], due to both internal and external factors. It was further discovered that the internal factors (syllabic position, vowel following the variables and lexical item status) are more highly connected to the dependent variables than the social factors. On one hand, [z] is favoured by back vowels, while [ð] is favoured by front vowels. On the other hand, with the status of lexical items and syllable position, [s] is likely to occur frequently in borrowed words from Swahili and in initial syllable position, while [θ] is likely to occur in native Chasu words and in the middle and final syllable positions. Concerning external factors, educational attainment is a significant social factor that is related to a high proportion of variant [z] against [ð] as well as [s] over [θ], with highly educated speakers in the lead. While variant [z] is used by highly educated speakers regardless of their age and gender difference, for [s] the CART-tree model showed an interaction between factor groups, where men who are highly educated from young and middle age groups use [s] more frequently than their counterparts and vice

versa. It is only the education level factor which appeared to influence both variables in this study, though it interacted highly with other factors, especially the internal ones.

Though variables (z) and (s) came into the Asu community through mingling of the northern and southern languages, those attaining a reasonable level of education will have extensive exposure to the standard variants through Swahili. Youngsters are more susceptible to the current need for education than any other group. Essentially, insufficient land in the rural villages and the changes in global economy make them remain in school, the only option as long as they can afford it. While in this study gender does not show a significant correlation to variable (z), it is, contrarily, a significant factor in relation to variable (s) of the same speech community. Men lead in the use of standard variant [s] over [θ].

Another important point is that, in Chasu society, social class does not clearly reflect in the variation of the variables for this study. This could be due to the fact that VARBRUL had insufficient cells to prove it as a result of unequal representation of informants i.e. due to factor interaction or the fact that, in rural African communities, social class is not yet a factor, which could affect the phonological variation and change as it happens in the Western urban communities.

Generally, the analysis in this chapter has shown inconsistent patterns of social factors in relation to phonological variation. Where social factors appear to be significant, they are highly interacted with internal factors, leading to the conclusion that, in this study, the structural constraints are stronger than social factors, as the data has proved through CART-tree analysis. Social factors work only subject to internal factors.

CHAPTER SIX
SUMMARY AND RECOMMENDATION

6.0 Introduction

This chapter presents summary and concluding remarks of the study and recommendations for further research.

6.1 Summary

This study was an investigation and analysis of sociolinguistic variation in rural African communities, among the individuals of Same District in Kilimanjaro Tanzania. The primary informants featured in this study are people who were born, raised and did at least their primary schooling within Myamba ward, hence speaking Southern Chasu dialect, widely spoken in the southern part of Same District. In the Chasu speech community, people can be classified with respect to the rest of their society by quantified characteristics like education, occupation, age and sex. Most of the people are subsistence farmers who, together with a few pastoralists constitute a lower class group of people with primary school education or no formal school at all. Another group, of middle class people is made up of educated professionals and businessmen and women.

The primary objectives were to examine how social factors such as age, gender, education, social class and style control the phonological variation of variables (s) and (z) within Chasu, as well as in the context of lexical borrowing and code-switching among the multilingual Chasu community. The investigations had to be measured against the sociolinguistic variation of various speech communities in Western urban societies.

The study was preceded by the pilot study work, which helped to determine the methods of gathering data, categorizing the linguistic and social variables and the effects of language contact given that Chasu is a multilingual community. Based on the

objectives of the study and the initial findings from the pilot study work, the Labovian model of sociolinguistic interview was the major technique employed in data collection during the main research work. In addition, techniques such as wordlist, participatory observation and questionnaires were employed to supplement data elicitation hand-in-hand with sociolinguistic interviews.

The data was recorded and transcribed for analysis and interpretation based on the objectives of the study. Concerning sociolinguistic variation, the focal point was phonological variables (s) and (z) with their variants [s] against [θ] as well as [z] over [ð] respectively. The analysis adopted the quantitative paradigm, which allows empirical examination of the model, which predicts the relationship between language variation and speaker's variables such as gender, age, social class, and education level and speech style. Thus, after transcription, the data was analysed quantitatively through the multivariate analysis of VARBRUL program and Rbrul analysis, particularly CART trees. Through this program models with significant factors which favoured the standard variants were selected against the non-standard; the insignificant factors were eliminated.

The study of language contact in this context involved three languages, Chasu being a host language, with English and Swahili mainly being donor languages. Concerning lexical borrowing, the data in this study has shown the frequent usage of loan words. Two types of loan forms, the core and non-core borrowed lexemes, were discussed. It was found that, except for the numerical forms which are borrowed most often, nouns and verbs from both core and non-core are the most frequently borrowed and used lexical categories. However, non-core forms are the most borrowed forms, as they tend to be associated with new borrowed culture. Likewise, in code-switching, where the Matrix Language Frame model of Myers-Scotton was employed, it was established that Chasu speakers code-switched to Swahili and English from word level category, phrases within clauses and sentences or at the sentential boundary.

In the study of phonological variation in monolingual speech, various social factors work together to shape the social identities of individuals, hence affecting their

language use. Likewise in multilingual language variation, code alternation is also affected by various social factors. Since Blom and Gumperz's 1972 study of code-switching in Norway, it has been accepted that alternation between codes is not random, but has social significance. Code-switching and lexical borrowing are more about constructing identity and social meanings associated with language use among Chasu speakers.

In Western societies, studies have consistently reported women to have a higher tendency of using greater proportions of standard variants than men of the same class within the same speech style. This pattern of gender differentiation has been consistent to the extent that many sociolinguists have considered it as a fundamental and universal principle of gender-based difference (Fasold 1990, Chambers 1992). In this study, gender was a significant factor in variable (s) but had no effect in variable (z). In variable (s), men favoured the standard variant [s] against [θ]. So there was no consistent pattern of gender differentiation in phonological variation. Concerning gender in relation to lexical borrowing and code-switching, women, especially the highly educated, use borrowed forms and code-switched to Swahili and English more frequently than men. Unlike in phonological variation, where there is no consistent pattern of variation in relation to gender, in the context of multilingual language variation, there is consistency. Women compete within and across their group. This authenticates the conclusion of Tehran (1978:73) in gender studies as he asserts that "any correlation of sex with linguistic variations should be considered more suggestive than definitive". This implies that male/female differences in language seem to be intimately involved in the mechanism of linguistic change, but also that either male or female can be a linguistically innovative gender (Coates 1986:151).

Concerning age factor in this study, young and middle-aged speakers used variant [s] against [θ] more frequently than middle and old-aged speakers, while age difference had no significant effect in variable (z). When age difference was associated with education level, highly educated young and middle-aged speakers were leading in using standard variant [s]. Likewise in the context of language contact, young and middle-aged

speakers used lexical borrowings and code-switched more frequently compared to the old-aged speakers, who maintained various features of the dialect.

In this study, style and social class had no significant effect on either phonological variable despite the fact that these are very important social factors in the Western world, where industrial development has set the society into a stabilized social hierarchy. On the other hand, with variation in multilingual context social class had an impact in regulating linguistic phenomena of language contact. Middle-class speakers code-switched and used borrowed forms more frequently than the lower-class speakers. It is important to concede that, in Chasu society, switching codes is not associated with producing stigmatized forms of speech, because the ELs have a higher social status than the ML, hence individuals value the use of features from a foreign language.

In relation to education level, the data has shown that the higher the level of educational attainment, the greater the chance of using standard variants. Likewise, highly educated speakers code-switched and used borrowed forms more frequently than less educated speakers. This suggests that code-switching is adopted mainly by individuals who have a certain degree of exposure to high education especially with switching to English. High educational attainment is the only social factor which had a consistent pattern in these linguistic phenomena, both in the phonological variation and in the context of multilingual effects. However, in the context of phonological variation, educational factors are significant only in interaction with other factors, especially the structural constraints of Chasu words.

Concerning internal factors, the data has shown that the syllable position, vowels following the variants and the status of the lexical item (i.e. whether borrowed from Swahili or a native Chasu word) are the most significant factors influencing the phonological variables (z) and (s) in Chasu community. Social factors are significant only when attached to internal factors.

From these findings, we can conclude that, together with other factors which may lead to phonological variation and change or the use of loan-words and code-switching, educational attainment is the only external factor which is highly associated with these linguistic phenomena, more so than any social factors in rural Chasu community. Individuals who are able to access education are exposed to the use of Swahili and English, and they borrow and code-switch; consequently, they become authors for language change. They also end up being sensitive to stereotyped variants; as a result, they consistently use prestigious variants. On the other hand, less educated speakers maintain traditions and linguistic features of the dialect.

6.2 Recommendations

This study has several recommendations for future research.

The study of how sociolinguistic variation and change signals social stratification is not a common area of research in rural African communities particularly in Tanzania. It was mentioned in chapter one that, together with variables (s) and (z) found in Southern Chasu dialect, there are other pairs of phonological variables such as post-alveolar fricative /ž/ and palatal plosive /ɟ/, velar fricative /ɣ/and velar plosive /g/, pre-nasalised consonants /mb/ and /mp/, /nd/ and /nt/, /nð/ and /nz/. The phonological difference in these variables is related to dialectics and their historical background. Based on this research, there is a great need for a broad empirical research in the Southern Chasu community to find out how social stratification regulates the variation of these variables in contemporary Chasu communities. The study should include a large sample of data, which will avail scientific and reasonable conclusions. It can involve a few factors which will help to reduce the interaction between factors, as Goldvarb appears to fail to handle such problems; otherwise Rbrul has to be employed from the beginning. The study can go beyond phonological variations to include other sections like semantic variation, especially with Swahili language in Tanzania.

Concerning the effects of language contact, it is evident that there are studies concerning language shift from ECLs to Swahili. However more, research could be done

on the structural effects of Swahili on ECLs and how social stratification regulates these structural changes in these languages. The study can be extended to include how the English language has affected both Swahili and ECLs in terms of lexicon and structural forms, especially as English is currently used globally as a disseminating agent of culture and scientific information.

REFERENCES

Abd-el- Jawad, H. 1981. Lexical and phonological variation in spoken Arabic in Amman, Ph.D. dissertation, University of Pennsylvania.

Abd-el- Jawad, H. 1987. Cross-dialectical variation in Arabic: Competing prestigious forms. *Language in Society.* 16:359-368.

Abdulaziz-Mkilifi, Mohamed, H. 1972. Triglossia and Swahili-English Bilingualism. *Language in Society.* 1:197-213.

Albo, X. 1970. *Social constraints on Cochabamba Quechue:* Ph.D. dissertation, Cornell University, Latin American Studies Program, Dissertation series No. 19.

Alfredo Ardila and Eliane Ramos. 2007. Bilingualism in the Contemporary World. *In Speech and Language Disorder*. 1-19. New York:Nova Science Publishers, Inc.

Alidou, Hassan. 2004. *Medium of Instruction in Post-colonial Africa*: *Medium of Instruction Policies.* London: Lawrence Erlbaum Associates.

Appel, R. and Muysken Pieter. 1987. *Language Contact and Bilingualism*. London: Edward Arnold.

Ash, Sharon. 2002. Social Class. In Chambers, J. K., P. Trudgill and N. Schilling-Estes (eds). *The Handbook of Language Variation and Change*. 402-422. Oxford: Blackwell.

Baayen, Harald R. 2008. Analysing *Linguistic Data: A Practical Introduction to Statistics using R.* Cambridge UK: Cambridge University Press.

Bayley, Robert. 2002. The quantitative paradigm. In Chambers J.K., P. Trudgill and N. Schilling-Estes (eds). *The Handbook of Language Variation and Change*. 117-139. Oxford: Blackwell.

Beebe, Leslie, M. 1997. The Influence of the listener on code-switching. *Language Learning.* 27:331-9.

Berdan, Robert. 1996. Disentangling Language acquisition from Language Variation. In Robert Bayley and Dennis, R. Preston (eds). *Second Language acquisition and Linguistic Variation.* 203-44. Amsterdam: John Benjamins.

Besha, Ruth Mfumbwa. 1989. *A study of Tense and Aspect in Shambala: Language and Dialect Studies in East Africa*. Berlin: Dietrich Reimer Verlag.

Blom, Jan-Petter and John J. Gumperz. 1972. Social Meaning in Linguistic Structures: Code-switching in Norway. In J.J. Gumperz and D. Hymes (eds.) *Directions in Sociolinguistics: the Ethnography of Communication.* 407-435. New York: Holt, Rinehart, Winston.

Blommaert, Jan. 1996. Language & Nationalism Comparing Flanders and Tanzania. *Nations and Nationalism.* 2:235-256.

Blommaert, Jan. 1997. *State Ideology and Language: The Politics of Swahili in Tanzania*. Duisburg: LICCA.

Blommaert, Jan. 2005. Situating language rights: English and Swahili in Tanzania revisited. *Journal of Sociolinguistics.* 9:390-417.

Cameroon, Deborah and Jenifer Coates. 1988. Some problems in the sociolinguistic explanation of sex differences. In Deborah Cameroon and Jenifer Coates (eds). *Women in Their Speech Communities: New Perspectives on Language and Sex.* 13-26. London and New York: Longman.

Chambers, Jack, K. 1992a. Dialect Acquisition. *Language.* 68:673-705.

Chambers, Jack, K. 1992b. Linguistics correlates of gender and sex. *English World-Wide.*13:173-218.

Chambers, Jack. K. 2002. Patterns of Variation including Change. In Chambers J.K., P. Trudgill and N. Schilling-Estes (eds). *The Handbook of Language Variation and Change*. 348-360. Oxford: Blackwell.

Chambers, Jack. K. 2003. *Sociolinguistic Theory.* (2nd Ed.) Oxford: Basil Blackwell.

Chambers, Jack. K. and Margaret Hardwick. 1986. Comparative sociolinguistics of a sound change in Canadian English. *English World –Wide.* 13:173-218.

Cheshire, Jenny. 1999. Spoken Standard English. In A.R. Bex and R.J Watts (eds). *Standard English: The Continuing Debate.* 129-45. London: Routledge.

Cheshire, Jenny. 2002. Sex and Gender in Variationist Research: In Chambers J.K., P. Trudgill and N. Schilling-Estes (eds). *The Handbook of Language Variation and Change*. 423-443. Oxford: Blackwell.

Coates, Jennifer. 1986. *Women, Men and Language.* London: Longman

Coulmas, Florian. 1989*: Language Adaptation*. Cambridge: CUP.

Coupland, N. 1980. Style Shifting in Cardiff work- setting. *Language in Society.* 9:1-12.

Deuchar, Margaret. 1988. A pragmatic account of women's use of standard speech. In Jennifer Coates and Deborah Cameroon (eds). *Women in Their Speech Communities: New Perspectives on Language and Sex*. 27-32. London and New York: Longman.

Dewalt, Kathleen, M. & Billie Dewalt. 2002. *Participant observation: A Guide for Fieldworkers.* Walnut Creek (CA): AltaMira Press.

Di Sciullo et al 1986. Government and Code-mixing. *Journal of Linguistics.* 22:1-24.

Docherty Gerry et al 1997. Descriptive adequacy in phonology: a Variationist Perspective. *Linguistics.* 33:275-310.

Dorian, Nancy.C. 1981. *Language Death: The life and Circle of a Scottish Gaelic Dialect*. Philadelphia: University of Pennsylvania Press.

Eckert, Penelope. 1988. Sound change and adolescent social structure. *Language in Society.* 7 :183-207.

Eckert, Penelope. 1989. The whole woman: Sex and Gender difference in variation. *Language Variation and Change*. 1:245-67.

Eckert, Penelope. 1997. Age as a Sociolinguistic Variable. In F. Coulmas (ed). *The Handbook of Sociolinguistics*. 151-167. Oxford: Blackwell.

Eckert, Penelope. 2000. *Linguistic Variation as a Social Practice*. Oxford: Blackwell.

Eckert, Penelope and McConnell-Ginet. 1992. Think practically and look locally: Language and gender as a community-based practice. *Annual Review of Anthropology.* 21:461-90.

Fasold, Ralph. 1984. *The Sociolinguistics of Society*. Oxford: Basil Blackwell.

Fasold, Ralph. 1990. *The Sociolinguistics of Language.* Oxford: Blackwell.

Feagin, Crawford. 1979. *Variation and Change in Alabama English: a Sociolinguistic Study of the White Community.* Washington, DC: Georgetown University Press.

Feagin, Crawford. 2002. Entering the community: Fieldwork. In Chambers J.K., P. Trudgill and N. Schilling-Estes (eds). *The Handbook of Language Variation and Change*. 423-443. Oxford: Blackwell.

Ferguson, Charles. 1959. Diglossia. *Word.* 15:325-340.

Ferguson, Charles. 2003. Diglossia: In C.B. Paulston and G. R. Tucker (eds). *Sociolinguistics: The Essential Readings.* 345-358. Blackwell Publishing Ltd.

Finlayson Rosalie, Karen Calteaux and Carol Myers-Scotton. 1998. Orderly mixing and Accommodation in South African Code-Switching. *Journal of Sociolinguistics.* 2:395-420.

Fishman, Joshua, A. 1972. *Advances in the Sociology of Language* 2. The Hague: Mouton.

Galabawa, Justinian C.J. 2001. Development and Issues Regarding Universal Education. (UPE) in Tanzania. ADEA Biennial Meeting. October, 7-11, 2001.

Gardener-Chloros, Penelope. 1991. *Language Selection and Switching in Strasbourg*. Oxford: Clarendon Press.

Gardener-Chloros, Penelope. 1997. Code-switching: Language selection in three Strasbourg Department Stores. In Coupland N. and Adam J. *Sociolinguistics: A reader and Course book.* 361-375. London: Macmillan Press.

Gordon, E. M. 1997. Sex, speech and stereotypes: why women's speech is closer to the standard. *Language in Society*.15:23-52.

Gordon, Raymond G. Jr. 2005 (ed). Summer Institute of Languages Ethnologue. http://www.ethnologue.com/ (accessed Nov.2, 2006).

Grimes, Barbara, F. (ed) 1996. *Ethnologue: Languages of the world*. Summer Institute of Linguistics, 13th Ed.

Gumperz, J.J. 1969. 'Cognitive Aspect of Bilingual communication', working paper no.28, Language behaviour Research Laboratory: Berkeley, Calif.: University of California Press.

Guthrie, Malcom. 1948. *The Classification of the Bantu Languages*. London: International African Institute.

Guy, Gregory. 1991. Contextual conditioning in variable lexical phonology. *Language Variation and Change.* 3:223-239.

Haeri, Niloofar. 1997. *The Sociolinguistics Market of Cairo Gender Class and Education.* London and New York: Kegan Paul International.

Haugen, Einar. 1956. *Bilingualism in the Americas.* A Bibliography and Research guide. Montgomery: University of Alabama Press.

Haugen, Einar. 1972. *The Ecology of Language.* Essays by Einar Haugen. California: Stanford University Press.

Hibiya, J. 1988. A Quantitative study of Tokyo Japanese, University of Pennsylvania Dissertation.

Hockett, C. F. 1950. Age-grading and Linguistic continuity. *Language.* 41:184-204

Hodge, R. W., P. M. Siegel and P. M. Ross. 1964. Occupational prestige in the United States, 1925-1963. *American Journal of Sociology*. 70:286-302.

Hoffmann Christian Thomas. 2006. Corpora and Introspection as corroborating evidence: The case of preposition placement in English relative clauses. *Corpus Linguistics and Linguistic Theory.* 2:165-195.

Holmes, Janet. 1997. *Women, Men and Language*. London: Longman.

Horvath, Barbara. 1985. *Variation in Australian English: The Sociolects of Sydney*. Cambridge: Cambridge University Press.

Höschele, Stefan. 2005. Christian Remnant - African Folk Church: The History of Seventh-Day Adventism in Tanzania, 1903-1980. Ph.D. dissertation, University of Malawi.

Hudson, Richard A. 1980. *Sociolinguistics.* Cambridge: Cambridge University Press.

James, Deborah. 1996. Women, men, and the prestige speech forms: a critical review. In Victoria Bergvall, Janet Bing, Alice Freed (eds). *Rethinking Language and Gender Research: Theory and Practice*. 98-125. London and New York: Longman

Johnson, Daniel Ezra. 2009. Getting of the Goldvarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass.* 3:359-383.

Kagaya, Ryohei. 1989. *A Classified Vocabulary of the Pare Language.* Tokyo: Institute for study of Languages and cultures of Asia and Africa.

Kagaya, Ryohei. 1989. A study of Tonal System of the Gonja Verbs and Nouns- A dialect of Southern Pare Language: In *Studies in Tanzanian Languages.* Tokyo: Institute for study of Languages and cultures of Asia and Africa.

Kagaya, Ryohei. 1989. *A* study in the Tonal system of the Mbaga Dialect of Southern Pare Language: In *Studies in Tanzanian Languages.* Tokyo: Institute for study of Languages and cultures of Asia and Africa.

Kiango John, G. 2005. Tanzania's Historical Contribution to the Recognition and Promotion of Swahili. *Africa and Asia.* 5:157-166.

Kimambo Isaria, Ndelahiyosa. 1969. Political history of the Pare of Tanzania. 1500-1900. Ph.D. Thesis: University of Dar es Salaam.

Kimura, Doreen. 1987. Are men's and women's brains really different? . *Canadian Psychology.* 28:133-47.

Kothari, C.R. 1990. *Research Methodology: Methods and Techniques.* Vishwaprakashan: New Delhi.

Kotz, Ernst. 1909. Grammatik Des Chasu in Deutsch Ost-Afrika: Pare Gabirge. In Mreta, Abel 1998. *An Analysis of Tense and Aspect in Chasu*. Their Form and Meaning in the Affirmative Constructions. (*Beitrage zur Afrikanistik*; 10). Zulg Bayreuth, University Dissertation.

Labov, William. 1963. The social motivation of a sound change. *Word*. 19:273-309.

Labov, William. 1966. *The Social Stratification of English in New York City*. Washington D.C.: Center for Applied Linguistics.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William.1972a. *Language in the Inner City*. Philadelphia: University of Pennsylvania.

Labov, William. 1975. *What is a linguistic fact?* Lisse: Peter de Ridder Press.

Labov, William. 1984. Intensity. GURT (Georgetown University Round Table) 84:43-70.

Labov, William. 1989. The child as a linguistic historian. *Language Variation and Change*. 1:85-98.

Labov, William.1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*. 2:205-254.

Labov, William. 1994. *Principles of Linguistic Change, I: Internal Factors*. Oxford: Blackwell.

Labov, William. 1998. The interaction of sex and class in linguistic change. In Jenny Cheshire and Peter Trudgill. *The Sociolinguistics Reader, Gender and Discourse. Vol*.2. 7-52. London: Rouutledge.

Labov, William. 2005. Quantitative reasoning in linguistics. To appear in Peter Trudgill Et al. (eds). *Sociolinguistics: an International Handbook of the Science of Language in Society*. 2nd ed. Berlin and New York: de Gruyter.http://www.ling.upenn.edu/~wlabov/Papers/QRL.pdf, 3.4.2007.

Labov, William. 2006. *The Atlas of North American Ennglish: Phonetics, Phonology, and Sound Change.* Berlin, New York: Mouton de Gruyter.

Lance, D.M. 1975. Spanish-English Code-switching. In E. Hernandez-Chaves, Cohen &Beltramo (eds). *El Lengua De Los Chicanos*. 138-153. Centre for AppliedLinguistics, Arlington, Virginia.

Legére Karsten. 2002. The "Languages of Tanzania" Project: Background, Resources and Perspectives. *Africa and Asia.* 2:168-186.

Lennig, Matthew. 1978. Acoustic measurement of linguistic change: The modern Paris vowel system. Ph.D. dissertation: University of Pennsylvania.

Lewis, M. Paul. (ed).2009. *Ethnologue: Languages of the World*: 16th edition. Dallas, Tex: SIL. International online version: http://www.ethnologue.com

Lieberson, Stanley. 1981. *Language Diversity and Language Contact: Essays by Stanley Lieberson.* Stanford: Stanford University Press.

Lopez, L. 1983. A Sociolinguistic analysis of /s/ variation in Honduran Spanish, Ph.D. dissertation: University of Minnesota.

Maccoby, Eleanor Emmons and Carol Nagy Jacklin. 1974. *The Psychology of Sex Differences.* Stanford, CA: Stanford University Press.

Mees, J. M. and Collins, B. 1999. Cardiff a real time study of Glottalization. In P. Foukes and G.J. Docherty (eds). *Urban Voice Accent: Studies in the British Isles.* 185-202. London: Arnold.

Mekacha, Rugatiri, D.K. 1993. *The Sociolinguistic Impact of Kiswahili on Ethnic Community Language in Tanzania*: *A case study of Ekinata*. Bayreuth: African Studies.

Mesthrie, Rajend. 1996. Review of J.K. Chamber's Sociolinguistic Theory. *English World Wide*. 16:295-299.

Michael, John. 1962. The construction of social class index. *Codebook for the Mobilization for Youth.* Appendix A. Mimeo. New York: Mobilization for Youth. Routledge, London and New York.

Migge, Bettina. 2002. Code-mixing, gender and social identities among Eastern Maroons. Paper presented to the 14th biennial conference of the Society for Caribbean Linguistics. The University of the West Indies, St Augustine, Trinidad and Tobago.

Migge, Bettina. 2007. Code-switching and Social identities in the Eastern Maroon Community of Suriname and French Guiana. *Journal of Sociolinguistics.* 11:53-73.

Milroy, Lesley. 1980. *Language and Social Networks.* Oxford: Basil Blackwell.

Milroy Lesley. 1987. *Observing and Analysing Natural Languages: A Critical Account of Sociolinguistic Method*. Oxford: Basil Blackwell.

Milroy Lesley. 1991. New Perspectives on the analysis of sex differentiation in Language. In Kingsley and Kwok (eds). *Sociolinguistics Today: International Perspectives.* 163-79. London: Routledge.

Milroy, James and Lesley Milroy. 1985. Linguistic Change, social network and speaker innovation. *Journal of Linguistics.* 21:339-384.

Milroy, James and Milroy Lesley. 1993. Mechanism of change in urban dialects: The role of class, social network and gender. *International Journal of Applied Linguistics.* 3: 57-77.

Milroy James and Lesley Milroy. 1997. Varieties and Variation. In F. Coulmas (ed). *The Handbook of Sociolinguistics*. 47-64: Oxford. Blackwell.

Milroy et al 1994. Glottal stops and Tyneside Glottalization: competing patterns of variation and change in British English. *Language Variation and Change*. 6:327-358.

Milroy, Lesley and Pieter, (eds). 1995. *One Speaker-Two languages: Cross disciplinary Perspectives on Code- Switching.* Cambridge: Cambridge University Press.

Milroy, Lesley and Matthew Gordon. 2003. *Sociolinguistics: Method and Interpretation*. Malden (MA): Blackwell Publishing Ltd.

Mochiwa Zacharia. S.M. 1979: The Impact of Kiswahili on Ethnic Languages: A case from Handeni District. Unpublished M.A. dissertation, University of Dar es Salaam.

Modaressi, Y. 1978. A Sociolinguistic investigation of modern Persian, Ph.D.dissertation: University of Kansas.

Mohamed Abdulaziz, H. 2003. The History of Language Policy in Africa with Reference to Language Choice in Education. *Towards a Multilingual Culture of Education.* UNESCO Institute of Education.

Mohamed A. Mohamed. 2001. *Modern Swahili Grammar.* Nairobi: East African Publisher Ltd.

Morale, H.L. 1986. Velarization of -/N/ in Puerto Rican Spanish: In D. Sankoff, (ed). *Diachrony and Diversity.* 105-113. Amsterdam and Philadelphia: John Benjamins.

Morrison, D.R. 1976. *Education and Politics in Africa: The Tanzania Case*. Heinemann: London.

Mous, Maarten. 2003. *The Making of a Mixed Language: The case of Ma'a/Mbugu.* Amsterdam and Philadelphia: John Benjamins Publishing Company.

Mreta Abel Yamwaka. 2000. The nature and effects of the Chasu-Kigweno contact in Mous Maarten and Kihore, Yared (eds). *Languages of Tanzania/Lugha za Tanzania.*

Mreta Abel, Yamwaka. 1998. *An Analysis of Tense and Aspect in Chasu*. Their form and meaning in the Affirmative Constructions (*Beitrage zur Afrikanistik*; 10). Zulg Bayreuth, University Dissertation.

Msanjila Yohana, Petro. 1999. The use of Kiswahili in Rural Areas and Its Implications for the future of Ethnic Languages in Tanzania: Unpublished Ph.D. Thesis: The University of Dar es Salaam.

Myers-Scotton, Carol. 1993a. *Social Motivations for Code-switching: evidence from Africa.* Oxford: Clarendon.

Myers-Scotton, Carol. 1993b. *Duelling Language: Grammatical Structure in Code-switching*. Oxford: Oxford University Press.

Myers-Scotton, Carol. 1995. A lexically based model of code switching: In Lesley Milroy and Pieter Muysken (eds). *One Speaker- Two Languages: Cross-Disciplinary: Perspective on Code-Switching.* 233-256. Cambridge: Cambridge University Press.

Myers-Scotton, Carol. 1997. Code-switching: In F. Coulmas (ed). *The Handbook of Sociolinguistics*. 217-236. Oxford: Blackwell.

Myers-Scotton, Carol. 2002. *Contact Linguistics*: Bilingual Encounter and Grammatical Outcomes. Oxford: Oxford University Press.

Myers-Scotton, Carol. 2006. *Multiple Voices: An Introduction to Bilingualism*. Malden: Blackwell.

Ndezi, P.E. 1979. Changing Patterns of Multilingualism: A case study of Igunga in Tabora Region. Unpublished M.A. Dissertation: University of Dar es Salaam.

O'Barr William, M. 1971. Multilingualism in a Rural Tanzanian Village. *Anthropological Linguistics*. 13:289-300.

Omari, C.K. 1991. *God and Worship in Traditional Asu Society*. The Makumira Publications, The Evangelical Lutheran Church of Tanzania; Verlag der Ev-Luth Mission Erlangen.

Paolillo John, C. 2002. *Analysing Linguistic Variation, Statistical Models and Methods*. CSLI Publications: California.

Poplack, Sankoff and Miller. 1988. The Social Correlates and Linguistic processes of Lexical borrowing and assimilation. *Linguistics*. 57:1-28.

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.1988. Language Status and language accommodation along a linguistic border. In Peter H. Lowenberg (ed). *Language Spread and Language Policy*: Issues, Implications and Case Studies. 90-118. Washington, DC: Georgetown University Press.

Poplack, Shana and Marjory Meechan. 1995. Patterns of Language Mixture: nominal structure in Wolof-French and Fongbe-French bilingual discourse. In Milroy, Lesley and Pieter Muysken (eds). *One Speaker-Two languages: Cross-Disciplinary Perspectives on Code Switching.* 199-232. Cambridge: Cambridge University Press.

Poplack, Shana. 2004. Code-switching: In Ammon, N. Dittmar, K.J. Mattheier & P. Trudgill (eds). *Soziolinguistik; International Handbook of theScience of Language*. 2nd ed. Berlin: Walter de Gruyter.

Pujolar, Joan. 2004. *Gender Heteroglossia and Power: A Sociolinguistic Study of Youth Culture.* Berlin: Walter de Gruyter.

Rampton, Ben. 1995. *Crossing: Language and Ethnicity among Adolescents*. London. Longman.

Rampton, Ben. 2007. Coming into Class; Language Migration and Class sensitivity in Britain. American Anthropological Association 10th Annual meeting 30 November, Washington DC.

Redouane, Rabia 2005. Linguistic Constraints on Code-switching and Code-mixing of Bilingual Moroccan-French Speakers in Canada. In James Cohen, Kara,T. McAlister, Kellie Rolstad and Jeff MacSwan: Proceedings of the 4th International Symposium on Bilingualism. Somerville, MA: Cascadilla Press. 1921-1923.

Rickford, J. R. 1986. The need for new approaches to social class analysis in Sociolinguistics. *Language and communication*. 6:215-21.

Roberts, J. 1993. The acquisition of new and vigorous sound changes by Philadelphia children. Paper presented at the 22nd Conference on New Ways of Analysing Variation: Ottawa.

Roberts, J. 1997a. Acquisition of Variable rules: a study of (-t, d) deletion in Pre- school children. *Journal of Child Language.* 24:351-372.

Roberts, J. 1997b. Hitting a moving target; Acquisition of Sound Change in Progress by Philadelphia Children. *Language Variation and Change.* 9:249-266.

Roberts, J. 2002. Child Language Variation. In Chambers J. K., P. Trudgill and N. Schilling-Estes (eds). *The Handbook of Language Variation and Change*. 333-348. Oxford: Blackwell.

Roberts, J. and Labov, W. 1992. Acquisition of a dialect. Paper presented at the Twenty First Conference on New Ways of Analysing Variation, Ann Arbor

Roberts, J. and Labov, W. 1995. Learning to talk Philadelphian. *Language Variation and Change.* 7:101-112.

Romaine, Suzanne. 1994. *Language in Society: An Introduction to Sociolinguistics.* Oxford: Oxford University Press.

Romaine, Suzanne. 1995. *Bilingualism*. (2nd ed.). Oxford: Basil Blackwell.

Romaine, Suzanne. 1999. *Communicating Gender*. Mahwah, NJ: Lawrence Erlbaum.

Roy-Campbell, Zalime M. 2001. *Empowerment Through Language*. Trenton, N.J. and Asmara, Eritrea: Africa World Press.

Rubagumya, Casmir, M. 1989. English Medium Instruction in Tanzania Secondary Schools: A Conflict of Aspiration and Achievements. *Journal of Multilingual and Multicultural Development.* 11:107-115.

Rubagumya, Casmir, M. (ed). 1990. *Language in Education in Africa: A Tanzanian Perspective.* Cleveland: Multilingual Matters.

Rubanza Yunus. I. 1979: The Relationship between Kiswahili and Other African Languages: The Case of Kihaya. Unpublished M.A. Dissertation: University of Dar es Salaam.

Sankoff, David and Shana Poplack. 1981. A Formal Grammar of Code-switching. *Papers in Linguistics.* 14:3-45.

Sankoff, David, Shana Poplack and Swathi Vanniarajan. 1990. The Case of the Nonce Loan in Tamil. *Language Variation and Change.* 2:71-101.

Sankoff, Gillian. 1980. *The Social Life of Language.* Philadelphia: University of Pennsylvania Press.

Sankoff, Gillian (ed). 1986. *Diachrony and Diversity*, Amsterdam and Philadelphia: John Benjamins.

Sigley, Robert J. 1997. Choosing your Relatives: Relative Clauses in New Zealand English. Unpublished doctoral dissertation: Victoria University of Wellington.

Sigley, Robert J. 2003. The importance of interaction effects. *Language Variation and Change.* 15: 227-253

Singler, John. 2009. Variationist sociolinguistics and the Niger-Congo languages of West Africa: The Non-state of the art. A paper presented in SAALA/LSSA conference in Cape Town. 28/09/2009.

Tagliamonte, Sali. 2006. *Analysing Sociolinguistic Variation.* Cambridge: Cambridge University Press.

Thomason Sarah, G. 2001. *Language Contact-An Introduction*. Washington DC: Georgetown University Press.

Trudgill, Peter. 1972. Sex, Covert Prestige and Linguistic Change in the Urban British English Norwich. *Language in Society*. 1:179-95.

Peter. 1974a. *Sociolinguistics: An Introduction*. Harmondsworth: Penguin Books Ltd.

Trudgill, Peter. 1974b. *The Social Differentiation of English in Norwich*. Cambridge: University Press.

Trudgill, Peter. 1983. *On Dialect: Social and Geographic Factors.* Oxford: Basil Blackwell.

Trudgill, Peter. 1986. *Dialects in Contact*. Oxford: Basil Blackwell.

Trudgill, Peter. 2002. Social Differentiation. In Chambers J.K., P. Trudgill and N. Schilling-Estes (eds). *The Handbook of Language Variation and Change*. 373-74. Oxford: Blackwell.

Vitale Anthony, J. 1981. *Swahili Syntax*. Dordrecht: Foris Publications.

Whiteley Wilfried, H. 1969. *Swahili: The Rise of a National Language.* London: Methuen.

Whiteley Wilfried, H. (ed). 1971. *Language Use and Social Change*. London: Oxford University Press.

Weinreich, Ulrich. 1953/1968. *Languages in Contact*. The Hague; Mouton.

Weinreich et al 1968. Empirical foundations for a theory of language change. In W. Lehman and Y. Malkiel (eds). *Directions for Historical Linguistics*. 95-188. Austin: University of Texas press.

Williams, Glyn. 1992. *Sociolinguistics: A Sociological Critique*. London and New York: Routledge.

Wolfram Walter, A. 1969. *A Sociolinguistic Description of Detroit Negro speech*. Washington, DC: Center for Applied Linguistics.

Wolfram Watt and Ralph W. Fasold. 1974. *The study of Social Dialects in the United States.* Englewood Cliffs, NJ: Prentice-Hall.

Yesse, Kanyuma. 2005. Social Motivations for Code-Switching among Members of Primary School Communities: A case study of Rabuor Ward, Tarime District. Unpublished M.A. Dissertation: University of Dar es Salaam.

Young, Richard and Robert Bayley. 1996. VARBRUL analysis for second language acquisition research. In Robert Bayley and Dennis R. Preston (eds). *Second Language acquisition and Linguistic variation*. 253-306. Amsterdam: John Benjamins.

APPENDICES

Appendix 1: Questionnaire and wordlist for sounds (s) and (z)

Appendix 2: A list of the languages of Tanzania by numbers of L1 speakers, 18 July 2007 (adapted from Languages of Tanzania Project (LOT))

Appendix 3: Data for (z) before recoding

Appendix 4: Data for (s) before recoding

Appendix 5: Data for (z) after recoding

Appendix 6: Data for (s) after recoding

Appendix 7: T-test for lexical borrowing and code-switching - Age groups

Appendix 8: T-test for lexical borrowing and code-switching - Gender

Appendix 9: T-test for lexical borrowing and code-switching – Educational levels

Appendix 10: T-test for lexical borrowing and code-switching – Social class

Appendix 11: Two way factor group comparisons for (z)

Appendix 12: Two way factor group comparisons for (s)