A GRAPHONOLOGICAL STUDY OF GĨKŨYŨ: AN OPTIMALITY APPROACH

KURIA PETER MBURU

C82/11301/08

A THESIS SUBMITTED TO THE SCHOOL OF HUMANITIES AND SOCIAL SCIENCES IN FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY, OF KENYATTA UNIVERSITY

MARCH 2021

DECLARATION

I confirm that this thesis is my original work and has not been presented in any other university. The thesis has been complemented by referenced works duly acknowledged. Where text, data, graphics pictures or tables have been borrowed from other works, including the internet, the sources are specifically accredited through referencing in accordance with anti-plagiarism regulations.

Signature……. Date:……

Name: **Kuria Peter Mburu**

Registration No: C82/11301/2008

This thesis has been submitted for review with our approval as university supervisors.

Signature: …… Date: ……

Dr Phyllis W. Mwangi

Department of Literature, Linguistics and Foreign Languages

Kenyatta University

Signature: …… Date: ……

Dr Kenneth Ngure

Department of Literature, Linguistics and Foreign Languages

Kenyatta University

DEDICATION

This work is dedicated to my children Waithĩra, Mbũrũ and Njoki.

ACKNOWLEGDEMENTS

This thesis is the culmination of the effort by a great many people, all of whom I am greatly indebted to.

My gratitude first goes to my Supervisors, Dr Phyllis W. Mwangi and Dr Kenneth Ngure. You provided guidance unflinchingly and unreservedly. Dr Mwangi helped me surmount hurdles, academic and otherwise, when often times, they threatened to derail progress. I learnt and for this, I will always be grateful.

Dr Ruth Ndung'u introduced me to the world of OT and passed me a compass to navigate it with. Accept my thanks.

I am grateful to the Université de Djibouti which provided me with an environment in which to exercise academic freedom and also resources with which I embarked on this journey. Neither can I forget to thank my colleagues there, especially Ms. E. Ayieko who demystified OT for me and prodded me on whenever I faltered.

Prof Wiese of Philipps University of Marburg, Institute for German Linguistics, came in handy when materials on OT and orthography proved scarce. He was kind enough to send me his original papers. Professor Bruce Peng of the State University of New York likewise sent me his original papers on Gĩkũyũ vowel harmony. To you Gentlemen of the Letters, I express my sincere gratitude.

The Department of Foreign Languages, Literature and Foreign Languages of Kenyatta University deserves special thanks. The Chair, Dr Purity Nthiga provided me with space to work from and offered me a job as a part-time lecturer. Dr Gerry Ayieko gave me invaluable advice on my work. Drs Gachara, Kanana, Kebeya, Gimode, Wangia, Njiiri, Ogutu, Nyamasyo, Itumo, Ikaria, Kiguru, Nandelenga, Mr Gecaga, Mr Omasaja and Ms Owili each encouraged me in their own way. Frashiah Gathogo ensured no glitches along the way while Sammy Bomba solved technical hitches for me. Mary always chipped in with kind words. To you all, I am indeed, grateful.

My immense gratitude also goes to my family: wife Hannah Wanjirũ and children, Juliet Waithĩra, Henry Mbũrũ and Trevy Njoki. They were there for me even when I neglected them and immersed myself in this research. My father and mother always asked me when I was graduating; thanks for keeping me on my toes.

Above all, I thank God for His sufficient grace, good health and for letting me see the completion of this work.

I do, however, hastily add that any errors in this work are wholly my responsibility.

TABLE OF CONTENTS

DECLARATION ii
DEDICATION iii
ACKNOWLEGDEMENTS iv
LIST OF TABLES xi
LIST OF FIGURES xiii
SYMBOLS AND ABBREVIATIONS xvi
ABSTRACT xviii
CHAPTER ONE 1
1.0. INTRODUCTION 1
1.1 Background to the Study 1
1.2 Statement of the Problem 8
1.3 Research Objectives 9
1.4 Research Questions 9
1.5 Research Assumptions 10
1.6 Justification for the Study 10
1.7 Scope and Limitations 13
1.8 Chapter Summary 14
CHAPTER TWO 16
LITERATURE REVIEW AND THEORETICAL FRAMEWORK 16
2.0 Introduction 16
2.1 Literature on Orthography Principles 16
2.1.1 Orthographical-Phonological Studies 20
2.1.2 Literature on OT in Orthography 23
2.1.3 Orthography Studies on Kenyan Bantu Languages 26
2.2. Theoretical Framework 27
2.2.1 Introduction 27
2.2.2 The Optimality Theory 27

2.2.3 The Correspondence Theory 33
2.2.4 Justification for use of the Optimality Theory 36
2.3 Chapter Summary 37
CHAPTER THREE 38
METHODOLOGY 38
3.0 Introduction 38
3.1 Research Design 38
3.2 Site of Study 38
3.3 Sampling Techniques and Sample Size 39
3.4 Research Instruments 40
3.4.1 Questionnaires 41
3.4.2 Transcripts from Tape Recordings 41
3.5 Data Collection Procedures 41
3.6 Data Presentation and Analysis 45
3.7 Ethical Considerations 46
3.8 Chapter Summary 47
CHAPTER FOUR 48
PRINCIPLES OF ORTHOGRAPHY, THE CONSTRAINTS AND THE GĨKŨYŨ GRAPHEMES 48
4.0 Introduction 48
4.1 Gĩkũyũ Consonant Phonemes and Graphemes 48
4.2 Gĩkũyũ Vocalic Phonemes and Graphemes 50
4.2.1 Gĩkũyũ Short vowels 50
4.2.2 Gĩkũyũ Long Vowels 54
4.3 The Data 55
4.3.1 The Plosives 67
4.3.2 The Gĩkũyũ Nasals 74
4.3.3 The Palatal Affricate /ⁿɟ/, <nj> 77
4.3.4 The Fricatives 79

4.3.5 The Alveolar Tap /ɾ/ 86
4.3.6 The Approximants 88
4.3.7 The Gĩkũyũ Short Vowels and Graphemes 89
4.3.8 Gĩkũyũ Long Vowels 94
4.3.9 Vowels in Concatenation 103
4.4 The Orthography Criteria and the OT Constraints 112
4.4.1 The Phonological Adequacy Criterion 114
4.4.2 The Simplicity Criterion and the Graphemes 121
4.4.3 Community Ownership Criterion and the Constraints 123
4.5 The Principles, Constraints and the Graphemes 124
4.5.1 The Plosives and the Affricate 125
4.5.1.1 The Voiced Bilabial Plosive /^{m}b/ 129
4.5.1.2 The Alveolar Plosive /^{n}d/, <nd> 130
4.5.1.3 The Velar Plosive /^{ŋ}g/, <ng> 131
4.5.1.4 The Palatal Affricate /^{ɲ}ɟ/, <nj> 132
4.5.2 The Nasals 133
4.5.2.1 The Palatal Nasal /ɲ/ Grapheme <ny> 136
4.5.2.2 The Velar Nasal /ŋ/ <ng'> 137
4.5.3 The Fricatives 138
4.5.3.1 The Voiced Bilabial Fricative /β/ 139
4.5.3.2 The Voiced Inter-Dental Fricative /ð/ 142
4.5.3.3 The Voiceless Palatal Fricative /ʃ/ 145
4.5.3.4 The Voiced Velar Fricative /ɣ/ 147
4.5.3.5 The Glottal Fricative /h/ 148
4.5.4 The Alveolar Tap /ɾ/ 150
4.5.5 The Front High [+ATR] Vowel /i/ 153
4.5.6 The Mid [+ATR] Vowel /e/ 159
4.5.7 The Front [-ATR] Vowel /ɛ/ 166
4.5.8 The [+Low] Vowel /a/ 170

4.5.9 The Back High [+ATR] Vowel /u/ 173
4.5.10 The Mid [+ATR] Vowel /o/ 177
4.5.11 The Back [-ATR] Vowel /ɔ/ 183
4.6 Chapter Summary 187
CHAPTER FIVE 190
UNDERREPRESENTATION AND OVERREPRESENTATION IN GĨKŨYŨ 190
5.0 Introduction 190
5.1 Orthographic Depth 190
5.2 Underrepresentation and overrepresentation 192
5.2.1 The Plosives and the Affricate 194
5.2.2 The Nasals 196
5.2.3 The Fricatives 197
5.2.4 The Alveolar Tap /ɾ/ 202
5.2.4 The Short Vowels 203
5.2.4.1 The Front [+ATR] Vowels /i/ and /e/ 203
5.2.4.2 The Front [-ATR] Vowel /ɛ/ 204
5.2.4.3 The Back [-ATR] Vowel /ɔ/ 205
5.2.4.4 The Back [+ATR] Vowels /o/ and /u/ 205
5.2.5 The Long Vowels 206
5.2.5.1 Long Vowels in Initial Word Position 206
5.2.5.2 Long Vowels in Medial and Final word positions 208
5.2.6 Vowels in Concatenation 208
5.3 Chapter Summary 210
CHAPTER SIX 212
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS 212
6.0 Introduction 212
6.1 Summary of Findings 212
6.2 Conclusions 219

6.3 Recommendations 220
REFERENCES 222
Appendix 1 234
QUESTIONNAIRE 234
Appendix II 238
WORDS WITH TARGET GRAPHEMES IN RESPECTIVE POSITIONS .. 238
Appendix III 241
KENYATTA UNIVERSITY GRADUATE SCHOOL RESEARCH AUTHORIZATION 241
Appendix IV 242
NACOSTI PERMIT 242

LIST OF TABLES

Table 2. 1: Relation of Dominance between Constraints 29
Table 2. 2: The Decisive Role of Candidate 2 30
Table 2. 3: Levels of Representation 32
Table 3. 1: Respondents by Gender and Dialect 40
Table 3. 2: Target Data 43
Table 3. 3: Targeted Concatenated Vocalic Graphemes 44
Table 4. 1: Gĩkũyũ Consonant Phoneme Inventory 49
Table 4. 2: Gĩkũyũ Consonant Grapheme Inventory 50
Table 4. 4: Reading Gĩkũyũ is Easy 57
Table 4. 5: Reading and Writing in English Easier than in Gĩkũyũ 58
Table 4. 6: Confusing Consonantal Graphemes 59
Table 4. 7: Confusing Vocalic Graphemes 62
Table 4. 8: Graphemes to be Excluded from the Inventory 64
Table 4. 9: Graphemes to be Included in the Graphemic Inventory 66
Table 4. 10: Respondents by dialect and sex against deviant plosive graphemes 69
Table 4. 11: Differing Phones for Prenasalised Phonemes produced by Respondents 70
Table 4. 12: The Gĩkũyũ Plosives 71
Table 4. 13: Respondents by Sex and Dialect against Deviant Nasal Graphemes 75
Table 4. 14: The Gĩkũyũ Nasals 76
Table 4. 15: Respondents Producing Variant Grapheme <j> 77
Table 4. 16: The Palatal Affricate 78
Table 4. 17: Respondents and Variant Fricative Graphemes 81
Table 4. 18: Respondents and Differing Fricative Phones 84
Table 4. 19: The Gĩkũyũ Fricatives 85
Table 4. 20: Respondents Producing Deviant <l> for <r> 86

Table 4. 21: Respondents Producing Phone [l] for the Tap /ɾ/..... 87
Table 4. 22: The Alveolar Tap 88
Table 4. 23: The Approximants..... 89
Table 4. 24: Deviant Graphemes, Number of Respondents by Dialect and Sex91
Table 4. 25: Short Vocalic Phonemes and Graphemes 93
Table 4.26: Variant Initial Long Graphemes, Number of Respondents by Dialect and Sex 97
Table 4.27: The Gĩkũyũ Long Vowels in the Initial Word Position 98
Table 4.28: The Gĩkũyũ Long Vowels in Medial and Final Positions..... 102
Table 4.29: Vowels in Concatenation 104
Table 4. 30: Front High [+ATR] /i/ Variant Graphemes against Number of Respondents 106
Table 4. 31: Front Mid [+ATR] /e/, Variant Graphemes against Number of Respondents 108
Table 5. 1: Concatenated Vowels and their Graphemic Representations……208

LIST OF FIGURES

Figure 4. 1: Gĩkũyũ Vowels 51
Figure 4. 2: The Seven Vowel system 52
Figure 4. 3: The Gĩkũyũ Short Vowels 53
Figure 4. 4: Gĩkũyũ Vocalic Graphemic Inventory 54
Figure 4. 5: Gĩkũyũ Long Vowels 54
Figure 4. 6: Gĩkũyũ Long Vocalic Graphemes 55

OPERATIONAL DEFINITION OF TERMS

Allographs: the various graphemes presented for a phoneme

Allophones: the different phonetic realizations of a phoneme

Constraint: a structural requirement that may either be satisfied or violated by an output form.

F_1: the first formant that is inversely related to vowel height

Faithfulness Constraints: constraints which require that outputs preserve the properties of the inputs.

Formant: a concentration of acoustic energy around a particular frequency in a speech wave

Grapheme: a unit of writing representing a phoneme.

Graphonology: study of the graphemic structure of a language based on the phonology of the said language.

Markedness Constraints: constraints which require that output forms meet some criteria of structural well-formedness.

Orthography: a system for representing a language in written form that encompasses more than the sounds.

Overrepresentation: a situation where an orthography uses more graphemes than there are phonemes.

Phoneme: the minimum unit of sound capable of distinguishing meaning

Script: a term that relates to graphic representation and which is not necessarily restricted to a single language

Underrepresentation: an instance where an orthography uses fewer graphemes than there are phonemes.

Writing system: a method of representing the sounds of a language by written or printed symbols

SYMBOLS AND ABBREVIATIONS

< >	:	notation of a grapheme
/ /	:	notation of a phoneme
/a:/	:	a long phoneme
*	:	mark showing violation of a constraint by a candidate form
*COMPLEX:		asterisk shows the negative quality of the constraint
!	:	mark showing fatal violation of a constraint by a candidate form
☞	:	winning candidate
ATR	:	Advanced Tongue Root
C	:	consonant
CONS	:	constraints
CT	:	Correspondence Theory
DEP	:	Constraint Family
F_1	:	the first formant (F1) in vowels is inversely related to vowel height, i.e. the higher the formant frequency, the lower the vowel height (and vice versa)
GEN	:	the operational component in OT
H-EVAL:		evaluator component that selects the most harmonic, most optional output
IDENT:		Constraint Family which ensure that correspondent segments are identical in features
In_k	:	linguistic input
(I, O)	:	input- output candidate pair

IPA :	International Phonetic Alphabet
LETTER-LETTER:	constraint which demands that a letter shape must not be changed.
MAX :	Constraint Family
MAX-IO:	where every segment of the input has a correspondent in the output.
MAX-BR:	where every segment of the base has a correspondent in the reduplicant
OT :	Optimality Theory
P*n* :	principle of orthography number *n*
$\mathcal{R}$:	correspondence relation between elements of each input and output level.
(S1, S2):	two strings candidate pair
SIMPLICITY:	constraint which demands that a grapheme is kept simple
SLC :	Sound Letter Correspondence
SPE :	Sound Patterns of English
UKLC :	United Kikuyu Language Committee
V :	vowel
VISIBILITY:	constraint requirement that salient features should be represented
VISIBIL$^{\text{V-LENGTH}}$:	constraints which demand that vowel length must be realized visually

ABSTRACT

This is a graphonological study of the Gĩkũyũ graphemic structure. The study sought to examine the extent to which the criteria for designing an orthography is observed, account for the graphemic structure of Gĩkũyũ using the Correspondence Theory, a sub-theory of Optimality Theory (OT), and evaluate the consequences of underrepresentation and overrepresentation of the orthography criteria on speakers and readers of Gĩkũyũ. Data was generated by respondents drawn from the five dialects of Gĩkũyũ, namely Gĩcũgũ, Mathĩra, Ndia, Northern Gĩkũyũ and Southern Gĩkũyũ. Each dialect produced six respondents, three males and three females, thus, a total of thirty respondents. They wrote translations of selected items from English to Gĩkũyũ to provide graphemic data. They then read their translations thus providing phonemic data. The phonemes and graphemes were pitted against the Principles of Orthography that underlie each criterion used in designing an orthography. To account for the graphemic structure, the data were analysed against OT constraints which were derived from the aforementioned principles. The findings include: the identification of hitherto unidentified vocalic concatenations; discovery of OT constraints that had not been used before; and, the establishment of Gĩkũyũ orthography as a deep orthography. The study recommends a phonetic study to establish the vowel space of Gĩkũyũ vowels, a revision of the Gĩkũyũ graphemic inventory so as to include the additional graphemes examined in this research, and, lastly, similar studies on Kenyan Bantu languages with a view of harmonizing their orthographies.

CHAPTER ONE

1.0. INTRODUCTION

This chapter presents information on the background to the study, statement of the research problem, research questions and objectives, and research assumptions. It also presents the justification for the study and the scope and limitation of the study on the Gĩkũyũ graphemic structure.

1.1 Background to the Study

This study aimed at evaluating the graphemic structure of Gĩkũyũ. In the 2019 census, Gĩkũyũ speakers numbered 8,148,668 (Kenya National Bureau of Statistics, 2019). The language is in the Central branch of the Bantu subgroup of the Niger-Congo language family, classified by Guthrie (1967) as E.51. There are five major varieties of Gĩkũyũ (Mwangi, Njoroge and Mose, 2013): Gĩcũgũ (spoken in Northern Kirinyaga), Mathĩra (spoken in Karatina), Ndia (in Southern Kirinyaga), Northern Gĩkũyũ (in Northern Murang'a, Nyeri), and Southern Gĩkũyũ (spoken in Kiambu, Southern Murang'a). The first Gĩkũyũ orthography was designed by Christian missionaries and the colonial administrators, with a more standardized orthography done by the United Kikuyu Language Committee (UKLC) in the 1940s (Kioko, Njoroge and Kuria, 2012).

The graphemic structure of a language entails the usage of graphemes to represent phonemes in that language. Graphemic structure also refers to the writing system which is a method of representing the sounds of a language by written or printed symbols. In this study, the term graphemic structure is used interchangeably with

the term orthography. Orthography generally refers to a system of representing a language in written form. It is encompassing and includes the placement of graphemes, word breaks, punctuation, diacritics, capitalization, and other features (Cahill & Karan, 2008). Besides, it is language specific.

The graphemic structure of a language aims at representing the phonological structure of that language with permanent, visible marks (Sampson, 1985). An orthography can be either phonemic (a shallow orthography) or non-phonemic (deep orthography) (Hasselbring, 2006). A phonemic orthography usually shows adherence to the one to one phoneme-grapheme principle where a single phoneme is represented by a single grapheme or a combination of graphemes. Examples in Gĩkũyũ are /t/ which is represented by < t > and /k / by < k >, and /ð/ by <th>. In a non-phonemic orthography, the relation between the graphemes and the phonemes is more opaque. There is usually no one-to-one correspondence between the grapheme and the phoneme. A single grapheme may have many different phonetic realizations, an example of which is <d> in English in the words ***d****iet*, *kicke****d***, and *bu****dd****y* which is realized as /d/, /t/, and /ɾ/ respectively. Conversely, a single phoneme may have different graphemic representations. A good illustration is /ʃ/ in English which may be variously represented as <sh> in the word '**sh**ip', <t> in the word 'sta**t**ion', <s> in the word '**s**ugar', and <ch> in the word '**ch**auffeur'. A study of the graphemic structure of a language is obviously based on the phonology of that language. This makes such a study

graphonological. The current research is, therefore, a graphonological study of the Gĩkũyũ language.

The evaluation of a writing system demands that one investigates the criteria used in setting the system so as to find out how well this has been achieved. There are some criteria that any orthography must meet. These include phonological adequacy, simplicity, areal appropriateness, community ownership, accuracy, consistency, convenience and harmonization (Bradley, 2003; Malone, 2004; Stark, 2010). These criteria do conflict. Languages rank them differently based on what is important to them, and sometimes violate them (Stark, 2010; Wiese, 2004; Baroni, 2016). That the criteria conflict and are ranked differently established the ground for examining the Gĩkũyũ orthography using the Optimality Theory.

The tenets of the Optimality Theory, specifically the Sub-theory of Correspondence (Prince & Smolensky, 1993) were applied in the analysis of data. The Correspondence Theory is the notion that there are inputs in a language which have equivalent outputs. The outputs have a certain order of preference which is determined by their harmony. The most harmonic outputs are those that have the least violations on some constraints, and those that have preserved the most properties of the input. Outputs with fewer violations are ranked higher on a hierarchical order while those with more violations are ranked lower. Thus, the

most optimal outputs will have observed the Faithfulness Requirement which dictates that the output preserves the properties of the input.

Phonological adequacy takes into account how phonological features of a language are represented graphemically. Prosodic features such as tone and stress, and phonological processes such as, labialization and palatalization may also be candidates for representation. To achieve this representation, principles / constraints are observed. Some of the principles are: use one letter for one phoneme; write as you speak; write as you would like to read; do not change an orthography; do not change the inventory of letters (Wiese, 2004; Baroni, 2016). The order of importance in which the constraints are ranked is language specific. Languages may also violate the principles. As a consequence, orthographies differ from language to language. The Optimality Theory was bound to be helpful in understanding the ranking and/or violation of the constraints.

Of interest in this research was the question of whether there might have been phonological features that were omitted or needlessly included in the designing of the Gĩkũyũ orthography, leading to overrepresentation or underrepresentation. To give examples, in English, phoneme /k/ is variously represented with <k>, <c>, <ck> and <q> which constitutes overrepresentation, or overdifferentiation (Karan, 2006). Representing two or more phonemes by a single symbol, for instance <e> for vowels /e/, and /ɛ/, like in the words ***e****ke* and ***e****gg*, constitutes underrepresentation, or underdifferentiation (Karan, 2006). This research

addresses the phenomena of both underrepresentation and overrepresentation in Gĩkũyũ.

The current orthography of the Gĩkũyũ language was proposed mainly by Armstrong (1967) under the aegis of the UKLC. She finds it unnecessary in Gĩkũyũ orthography to mark long vowels by double letters as advocated by the Institute of African Languages and Cultures (Armstrong, 1967). She argues that it 'would seem generally advisable to record as long, those vowels which are long in words pronounced in isolation' (p. 1). As it is, in written language, sounds are represented by individual graphemes but in spoken language, the sounds run into each other; the boundaries are often indistinct. As Kerr (1998, p. 44) notes, "Phonemes are often barely delineated at all in conversational language; even word boundaries are usually very comprehensively blurred". For instance, the name 'Njeri' (name of a person) may sometimes be vocalized as /ɟɛri/, /^{n}ɟɛri/ or /^{n}ɟɛɛri/ depending on such variables as stress, speaker or audience. How then is phonological adequacy achieved in such a case? How should such a word be represented in the orthography? These are issues of underrepresentation and overrepresentation. This study proposed to find out what Gĩkũyũ writers and readers do when they are confronted with words carrying vocalic sounds that can be read or written in more than one way.

The representation of two consecutive vowels in writing has to do with phonological adequacy. Armstrong (1967) observes that in some cases,

consecutive vowels form one syllable but in the majority of others, they are disyllabic. Thus, <iĩ>, < iɛ>, <ia>, <iɔ>, <iũ>, <iu>,<ĩi>, <ɛi>, <ai>, <ũi>, <ɔi>, <ĩɛ>, <ĩa>, <ĩɔ>, <ĩũ>, <ũĩ>, <oĩ>, <uĩ>, <ũɛ>, <uɛ>,<ũa>, <ua>, <uɔ>, and <uũ> are in some cases monosyllabic, and disyllabic in others. The choice of representation could be as a result of phonological processes such as gliding, coalescence, dissimilation or any other. The current study intended to find out if such sequences of vocalic sounds are represented by Gĩkũyũ writers as one, two or even more graphemes, and why.

Kuria (2005) reports that there is need to investigate the presence of or lack of the homorganic nasal element that Armstrong (1967) says should always accompany the phonemes /b/, /d/, /g/ and /dʒ/. This is out of his findings that authors and readers sometimes represent the nasal orthographically and at other times omit it without any justification either way. This research sought to determine whether there is need to graphemically represent the nasal segment.

On the same issue of phonological adequacy, it is also important to note that written language is more conservative than spoken language. Consequently, the graphemic system of a given language might not appear as representative of any of the dialects as a result of diachronic changes. This is true of a language like English which has a history of several centuries. On the other hand, the orthography of Gĩkũyũ is hardly a century old, having been proposed in 1947 (Armstrong, 1967). Some writers may be trying to represent changes that the

language might be undergoing. We, therefore, investigated any such changes and the extent to which writers have been trying to represent them orthographically as we examined the phonological adequacy criterion. Currently, Gĩkũyũ is said to have eighteen consonants. These are /^{m}b/ /β/ /m/ /ɲ/ /t/ /ð//^{n}d/ /r/ /n/ /ʃ/ /^{ɲ}j/ /k/ /^{ŋ}g/ /ŋ/ /w/ /h/ /j/ and /ɣ/ (Iribemwangi, 2012). Their graphemes are <mb>, <b>, <m>, <ny>, <t>, <th>, <nd>, <r>, <n>, <c>, <nj>, <k>, <ng>, <ng'>, <w>, <h>, <y> and <g>, respectively. It has seven vocalic phonemes /a/, /ɛ /, /i/, /e/, /ɔ/, /u/ and /o/ whose graphemic representations are <a>, <e>, < i >, < ĩ >, <o>, <u>, and <ũ>, respectively (Mwihaki, 1998).

The issue of community ownership is important in designing an orthography. Do all the speakers of the language feel that the orthography represents their phonology? Do some feel that there were omissions and others that some 'strange' phonological elements were included? In the design of Gĩkũyũ orthography, Armstrong (1967) relied heavily on the Southern Dialect and, more so, on the 'Kikuyu of Mr. Kenyatta' (Kabete Dialect). Obviously, this was to the detriment of the other dialects. Mutahi (1977, p. 9) notes that "the written word has done great harm to the small dialects", an implication that the small dialects were not fully represented in the orthography. Bradley (2003, p. 1) notes that 'orthography development … should involve the entire speech community'. In this study, we proposed to find out the attitude of Gĩkũyũ speakers towards their written language by posing the following pertinent questions: Do they feel that the

orthography accurately encodes their spoken language? Is the orthography adequate for their needs?

Another issue of major concern is orthographical simplicity. This has to do with readability and write-ability — the ease of reading and writing of the graphemes. Are the graphemes used for writing Gĩkũyũ simple enough to the native reader and speaker? Is the reader/ writer able to use them with ease or do they present unwarranted challenges? These are issues that this research sought to address.

Stark (2010) delineates some seventeen principles for an orthography development that she calls advanced. Some have to do with the phonological adequacy, others with community ownership, and yet others with simplicity. She emphasizes that these principles do not function independently, and that they are interrelated, though some may provide stronger arguments in specific situations. This research examined how these principles apply in the Gĩkũyũ orthography.

1.2 Statement of the Problem

The graphemic structure of a language represents the phonological structure of that language with permanent, visible marks. In order to achieve the representation, some criteria for orthographical development must be used. Such criteria include phonological adequacy, simplicity, and community ownership among others. The principles underlying the criteria are universal but violable. A language takes the principles and ranks them to produce its correct orthography.

Where the ranking of the principles has been observed, the orthography would be ideal and there would be no need to examine or even evaluate it. However, in the orthography of Gĩkũyũ, there exist inconsistencies, omissions and / or unnecessary inclusions. This creates the need to investigate how the orthographic criteria have been applied in setting the graphemic structure of the language.

This research, therefore, sought to explain and account for the discrepancies in the Gĩkũyũ graphemic structure and recommend a standard orthography of the language.

1.3 Research Objectives

To carry out this research, the following were our objectives:

1. To examine the extent to which the criteria for designing an orthography are observed in the orthography of Gĩkũyũ.
2. To account for the Gĩkũyũ graphemic structure using the Correspondence Theory of the Optimality Theory.
3. To evaluate the consequences of both the underrepresentation and overrepresentation of the orthographic criteria on the writers and readers of Gĩkũyũ.

1.4 Research Questions

This research aimed to answer the following questions:

1. To what extent are the criteria for designing an orthography observed in the orthography of Gĩkũyũ?
2. How can the Correspondence Theory of the Optimality Theory account for the Gĩkũyũ graphemic structure?
3. How do the underrepresentation and overrepresentation of orthographic criteria impact on the writers and readers of Gĩkũyũ?

1.5 Research Assumptions

This research proceeded from the following assumptions:

1. That certain criteria for orthographical design may have been flouted in the designing of Gĩkũyũ orthography.
2. That the Correspondence Theory of the Optimality Theory can account for the Gĩkũyũ graphemic structure.
3. That there is underrepresentation and overrepresentation of the orthographic criteria and they impact on Gĩkũyũ writers and readers.

1.6 Justification for the Study

Justification for this study exists in areas such as education, linguistics and communication.

The policy on language in education in Kenya is that in the first three years of primary school, the language of study and of instruction is the language of the school's catchment area: that is, the vernacular spoken in the environs of the primary school (Gachathi Report, 1976). Indeed, the Basic Education Curriculum

Framework emphasises that vernacular is the language to be used in Pre-primary school, and in Literacy and Indigineous Language activities in Grades 1-3 (K.I.C.D., 2017, p. 30). This is in line with the conviction that in learning in mother tongue, the learner develops cognitive gains and the development of his unique characteristics, his social awareness and social identity are heightened (Ministry of Education, 2002). Researchers have noted that the curriculum for mother tongue instruction was vague and not language specific and that the teachers were not trained for teaching in mother tongue (Okombo, 2010). There is further observation that there is "need to develop a mother tongue curriculum" (Okombo, 2010, p. 15). Since this study endeavours to suggest an ideal and uniform orthography, the findings and suggestions should be useful in developing a Gĩkũyũ curriculum and in the preparation of materials in Gĩkũyũ that are to be used in school. They may also be useful in the training of teachers in mother tongue teaching.

There are institutes of African languages that are teaching languages such as Gĩkũyũ and Dholuo. There are also language centres, for example, the ACK Language and Orientation school, where Gĩkũyũ is taught and learnt. Findings from this research would be beneficial to them with regard to the orthography to use when designing their teaching and learning materials.

There has been an explosion in communication, occasioned by the use of e-mail and text messaging. Though there are no documented studies on the use of this

technology in Gĩkũyũ, we hope that findings from this research would be useful for those who may want to use Gĩkũyũ in their e-mailing and in text messaging as well as to researchers in the field of communication. Additionally, the findings would also be beneficial to those who write, read and publish in Gĩkũyũ.

Debate on what the standard orthography of Gĩkũyũ should be is still ongoing (Kuria, 2005; Githiora, 2003). The findings from this research constitute our contribution to the debate and could be used to either revise or further develop the orthography.

There are efforts to harmonize the orthographies of the Kenyan Bantu languages of Kikamba and Gĩkũyũ (Kioko et al., 2012; Mwangi et al., 2013). Scholars are also making efforts towards the harmonization of other African languages (Prah, 2003; Banda, 2002). Findings from this research constitute our contribution to these endeavours.

Many studies have been carried out on the Gĩkũyũ language using different linguistic theories. These include the Autosegmental Phonology Theory (Mwihaki, 1998), Phonological Recodability and Linguistic Mental Representation (Kuria, 2005), the Merger Theory and the Incorporation Theory (Mwangi, 2001), and the Minimalist Theory (Waweru, 2011), among others. As far as the researcher can ascertain, no study on Gĩkũyũ orthography has been done using the Optimality Theory. Indeed, Wiese (2004) notes that there are very few contributions to the analysis of writing systems and orthography within the OT

framework. This is a study where the universal claims of OT are applied in an evaluation of the Gĩkũyũ orthography.

1.7 Scope and Limitations

We acknowledge the many phonological studies that have been done on Gĩkũyũ. However, it has been noted that authors in the language orthographically represent the same phonemes in different ways. Indeed, Githiora, 2003, p. 88 notes that "publications... continue to reflect haphazard, individual decisions rather than systematic, collective ones". In another language, for example, English, these would be deemed as errors. We sought a graphonological explanation to this phenomenon and used the Correspondence Theory of the OT theory, which is "adjudged as most current, efficient, viable and dependable because it is seen to have more explanatory force' (Omachonu, 2008, p. 95). In this qualitative study, we limit ourselves to strictly evaluating the consonantal and vocalic sounds and their graphemic representations in Gĩkũyũ.

As noted elsewhere, there already exists an orthography of Gĩkũyũ. Although initially we were limited to the current orthography, we had no intention of confining ourselves to only those graphemes. This study will have ramifications on the orthography of Gĩkũyũ and findings may be used for revision of the orthography. We were, therefore, not constrained to the conventional graphemes used in Gĩkũyũ orthography and may indeed, suggest others.

There are several studies on the dialects of Gĩkũyũ with different findings. In this study we only refer to these findings in regard to orthography-phonology issues. The data for our study is collected from the five dialects of Gĩkũyũ language.

In an orthography, tone marking may remove ease of writing and reading (Kioko, 2002; Seifart, 2006). A compromise on what to use may be necessary. To address tone marking in orthography, we would have to study tone structures, melodies, spread and shift, downstep and upstep, among other concerns, a feat that would considerably widen our scope. Consequently, in this study, the representation of tone orthographically was not addressed.

The research needed respondents who were proficient in reading and writing in Gĩkũyũ. However, there is no established assessment of such proficiency and the educational syllabus is silent on the same. Thus, the researcher had to rely on the self confessions of the respondents as to their ability to read and write in the language. To overcome this limitation, the researcher had to insist on this ability before the subject could take part in the study.

1.8 Chapter Summary

In this chapter, we have looked at the background to the current study in which we introduced the Gĩkũyũ language and its dialects, discussed what the graphemic structure of a language entails, and highlighted some issues with the Gĩkũyũ orthography. We have then stated the research problem and outlined the research objectives, the questions and the assumptions. Next, we have presented the

justification for the study and indicated that our findings could benefit Gĩkũyũ curriculum developers, teachers and learners, and contribute to communication through the internet, among other benefits. Lastly, we have provided the scope and limitations of this study. In the following chapter, we review the relevant literature and discuss the theoretical framework used in the study.

CHAPTER TWO

LITERATURE REVIEW AND THEORETICAL FRAMEWORK

2.0 Introduction

In this chapter, we first present literature on orthography principles followed by literature on orthographical-phonological studies. Next, comes literature on orthography studies done using OT, and lastly, literature on orthography studies of Kenyan Bantu languages. Under the theoretical framework, we first discuss the OT and then delve into the Correspondence Theory. Lastly, we present the justification for the use of the OT.

2.1 Literature on Orthography Principles

A study by Stark (2010) deals with the orthography design of the Kambari languages of Agwara, Auna and Salka found in Nigeria. She examines the foundations of orthography design, describes it and discusses the philosophy behind it. She emphasizes that the key factor in the development of an acceptable writing system is the principle of community participation. The author goes beyond the basic criteria of orthography development to develop a series of principles intended to guide orthography decisions in situations where the basic principles are inadequate. In our study, we compressed her principles into some three principles namely: phonological adequacy, community ownership and simplicity. We, nevertheless, differed from this research in that while the

principles are applied to the development and designing of an orthography, we applied them in the evaluation of an already existing orthography.

Bradley (2003), in advocating orthography development and reform for the Yi and Lisu languages of China, observes that there are competing and sometimes incompatible requirements for an orthography. These include phonological adequacy, simplicity, areal appropriateness and transfer to literacy in a natural language. Other requirements are appropriate technology, sufficient materials, and community ownership. He notes that many successful orthographies do not meet many of these criteria and that the overriding criterion is that the community must be happy with, and in control of its orthography. He adds that a variety of materials for teaching and using the orthography are essential. He also notes that orthographies become archaic through time but an established literary tradition often prevents reform. Although in this research we did not deal exclusively with orthography reform or development, the issues of phonological adequacy, simplicity and community ownership were investigated against a backdrop of the existing Gĩkũyũ orthography.

Malone (2004) discusses planning and implementing education-for-development programmes for adult speakers of minority languages. She describes some general guidelines that may be used in developing writing systems for languages. She notes that though there are no specific rules for developing a writing system, there are indeed general guidelines. One of the guidelines is that the writing system be

acceptable to mother tongue speakers of the language. To achieve this, she recommends that the opinions of mother tongue speakers on how their writing system should look like should be sought. Another consideration is that the writing system should represent the important features of the language accurately. She highlights the need for consistency in the representation of consonants, vowels and other features such as stress and tone. In the representation of sounds not found in other related or major languages, she notes the need for using symbols that are not in use in those languages, or borrowing a symbol from a neighbouring language, or even creating their own symbol. Though the objective of her study is different from ours, the issues of acceptability of a writing system, and phonological adequacy were of interest to our study. We also examined consistency in the representation of consonants and vowels but did not deal with stress and tone.

Guerin (2008) writes on an undocumented and endangered Oceanic language called Mavea, spoken on the Island of Mavea. She describes the orthography principles she employed in writing down stories she had recorded, as well as in compiling a dictionary in the language. She groups the principles into four categories: technological usability, social and cultural acceptability, psycholinguistic acceptability, and the function of the orthography. Her social and cultural acceptability principle is about the linguistic community's and/or the Ministry of Education's approval of the orthography. Our research takes the first part on the linguistic community's approval as a central point of study. Her

contention that the ‘best ethical practices in language documentation demand that the linguist respects the will of the community s/he works with’ (p. 63) is highly relevant. We investigated whether this was done in the designing of Gĩkũyũ orthography but we differed with the researcher in our aims. Hers were, in general, the application of orthography principles. Our aim was examining how those principles have been applied in a specific language, Gĩkũyũ in this case.

Another relevant study is that of Hasselbring (2006), who examines cross dialectal acceptance of written standards among the Leleni and the Likpakpaanl of Ghana. She presents literature concerning language standardization and the development of written languages, giving special attention to influence of language variation on standardization, standardization of previously written languages and language standardization that occurs without government involvement. She claims that speakers of languages for which a written standard has only recently been developed are usually more aware of, and have stronger feelings about differences between their spoken variety and the written standard. Speakers of each dialect feel that the written form of their language should match the way they speak. When materials in a newly developed standard do not match a speaker’s spoken dialect, negative attitudes about the standard may develop. This impacts their acceptance. In light of the fact that Gĩkũyũ orthography is relatively new and some dialects may have been neglected (Armstrong, 1967) in the making of the orthography, we found this researcher’s contentions relevant to our study. Issues of phonological adequacy and community ownership are

touched on here. We, nevertheless, differed from this research in that while she researches language standardization and language development, we were examining an extant orthography.

2.1.1 Orthographical-Phonological Studies

In an investigation on vowel length in Eastern Bantu orthographies, Stegen (2005) identifies phonemic length, vowel concatenation and compensatory length as being of considerable importance in deciding vowel length. He observes that underlying short vowels are lengthened in the environment of labialization, palatalization and prenasalisation. He also observes that phonemic length and vowel concatenation should typically be written with a double letter. He identifies several sources of vowel length as underlying representation which is the same as phonemic length, vowel concatenation, gliding plus compensatory lengthening, compensatory lengthening preceding a moraic nasal plus consonant, and penultimate vowel lengthening. We found his arguments about vowel length relevant as this is an issue in the orthography of Gĩkũyũ.

In a discussion on the relationship between orthography and phonology of an artificial language, Sasaki (2003, p. 96) observes that a ‘written language is governed by the writing system so that the readers will know a sound from a spelling or simply spellers will know a spelling from the sound’. He also observes that in a language, there might be hidden rules which do not represent actual sounds but provide visual cues for the sounds such as a double consonant letter

that marks the previous vowel as a short vowel in the words <latter> and <later> in English. In our research, we examined the orthography to see whether such claims could be true for Gĩkũyũ.

Sasaki (2003) further discusses how orthographic information in the lexicon is stored. He says that it is stored, first, on the basis of the correspondence between the grapheme-phoneme patterns. This correspondence can be consistent, inconsistent but predictable, or inconsistent and unpredictable. Secondly, the information is stored on the basis of the regularity of letters or letter sequences, that is, their frequency. In our study, we endeavoured to determine whether such correspondence and frequency are present in the Gĩkũyũ orthography.

Kioko (2002, p. 232) notes that "without the tone marking in the orthography, it becomes difficult to relate the spoken language to the written language since a major component of the spoken language is not represented in the writing system". Since tone marking may remove the ease of writing and reading (Seifart, 2006, p. 293), a compromise on what is rendered as the orthography has to be reached. Although we did not intend to study tone representation in orthography, we, nevertheless, wanted to find out which other features were compromised in settling for the current Gĩkũyũ orthography.

Baker (1997) observes that the phonological analysis of a language on which the orthography was originally based may have been faulty in some respects. One reason for this could be because an outside linguist failed to recognize some of the

phonemic differences. This could lead to non-representation, under-differentiation or overrepresentation of a feature. He notes that linguistic inadequacy can also arise from a deliberate decision to not represent a certain feature (for example, one with a heavy functional load such as tone in certain tone languages). In our study, we investigated whether non-representation, underrepresentation and overrepresentation are issues in Gĩkũyũ orthography.

Coulmas (1999) gives a cause for phonological inadequacy as the mismatch between the writing system and the language it is to serve. He explains that the abstract underlying principles of the writing system may not have been fully understood when it was first adopted for a given language. Consequently, the resulting orthography would prove to be inadequate. He gives the example of morphemic Chinese characters which worked well for non-inflecting languages but did not serve Korean or Japanese well. For them, something else was needed to represent their more complex multi-morphemic word structures. This research attempted to find out if there are such mismatches in Gĩkũyũ and what their causes are.

Baker (1980) notes that rigid adherence to the phonemic principle can result in a positive loss of relevant linguistic information. He gives one example of such loss as in the representation of the English past tense markers /t/ and /d/. Another example that he gives is plural markers /s/ /z/ /iz/ as used in signalling derivational relationships; in distinguishing homophones; and in differentiating

between lexical and grammatical morphemes that sound identical, such as the ‘s’ in plural and the ‘se’ singular in the words ‘pleas’ and ‘please’. He also gives the example of /ɵ / and / ð/ which may be represented with <th>. In our study, issues of phonemic adherence were examined to determine their presence and also their effect on the orthography of Gĩkũyũ.

2.1.2 Literature on OT in Orthography

Baroni (2016) analyses differences between orthographies from the Optimality Theory perspective. He notes that orthographies are best described as a combination of different principles which are often in conflict with one another. This makes OT best placed to analyse orthography. He proposes constraints that are grounded either in cognition or in phonetics. He further discusses vowel and consonant length in light of the different constraints. Lastly, he discusses Spelling Pronunciation and Eye Dialect. Spelling Pronunciation consists of the modification of the pronunciation of certain words based on how they are spelled. Eye Dialect leaves pronunciation unaffected but changes the spelling, normally to give the impression of non-standard speech. He notes that the two bring the phonological and the orthographic forms closer. The arguments presented in this work are highly relevant to our study. Constraint identification and ranking, spelling and pronunciation are issues that were investigated in this research.

Wiese (2004) applies OT to orthography principles on German orthography. He discusses a preference-oriented versus a rule-oriented systematic theory of

orthography. This is a theory of constraints and their interaction. He notes that writing systems follow general rules or principles and that there is an obvious functional motivation to these rules. He adds that optimization is not well understood because formal descriptions of orthographic regularities in terms of structure or generative graphemics do not normally include the concept of optimization. He justifies the use of OT as follows: firstly, that the notion of optimization is built into the theory itself; secondly, that the basic mechanism of evaluating parallel structures is particularly fit to deal with orthography; thirdly, that OT is able to deal with conflicts between preferences in a direct way. He discusses preferences in orthography, noting that they conflict, and that they are rarely true in an absolute sense. Using the Correspondence Theory, he examines the principles under 'Alignment' and 'Linearity' and presents some correspondence constraints. This study is highly relevant to our work. We apply some of its ideas about the inputs – outputs and its notion of correspondence to Gĩkũyũ orthography.

A related study is by Wiese and Song (2010) who apply OT principles to examine the relationship between the phonology and the orthography of both German and Korean. They identify language inputs as unpredictable letter shapes, underlying phonological forms, and lexical outputs. Faithfulness and Markedness constraints are identified as interacting in the study. Faithfulness Constraints of DEP, MAX and IDENT force the output to preserve the properties of the input in line with the Correspondence Theory (McCarthy and Prince, 1995). These Constraints are the

link between phonology and orthography. The Faithfulness Constraints are then divided into two: SOUND LETTER CORRESPONDENCE (SLC) constraints, and VISIBIL constraints. The first set of constraints relate phonemes to graphemes, while the second set require linguistic features such as tone and stress to be visualized. The researchers propose other constraints: the LETTER-LETTER constraint which demands that a letter shape must not be changed and the VISIBIL$^{\text{V-LENGTH}}$ which demands that vowel length must be realized visually. This study provides us with several constructs for our study. Firstly, we use their ideas of inputs to identify the inputs in our study. Secondly, we borrow and apply the ideas of the SLC constraints and the VISIBIL constraints, in addition to our own ideas.

Smith (2004) addresses the problems involved in applying OT to the study of orthography. He does this in the reconstruction of the phonology of Elamite, a language known only from written sources. With only a set of overt forms (orthographic symbols), he shows how a learner can derive both the underlying forms (phonology) and the set of OT constraints. He explains how the ranking of those constraints would lead to usage of OT's Lexical Optimization module to derive the underlying forms. Smith's ideas are highly relevant to our study. We use the same theory of ranking constituents to achieve the overt forms (orthography) from the underlying forms (phonology), though, in our study both forms are for an existing language.

2.1.3 Orthography Studies on Kenyan Bantu Languages

One of the studies done on a Kenyan Bantu language is by Wangia (2003). She discusses inconsistencies in the orthographic representation of the sounds of Lulogoli, a Luhyia dialect. She observes that there is no distinction between short and long vowels from the spelling of words; that vowel length, which is a significant feature in distinguishing meaning of words, is not marked. Vowel length could be marked by doubling the vowels. She also notes that there is arbitrary use of orthographic symbols representing consonantal sounds, undermining the predictability of the graphemes used. For example, phonemes /v/ and /b/ are interchangeably represented with <v> and <b>. This study sheds some light on discrepancies between the orthography and the phonology of a Kenyan Bantu language, just like Gĩkũyũ is.

Kioko et al. (2012) present a unified orthography for Bantu languages in Kenya. They give some general unified spelling rules for vowels, consonants and some morphemes. Just like in our study, they leave out the representation of tone orthographically. This study is important to us as it establishes the phonemic inventory of Gĩkũyũ and identifies some discrepancies with the orthography. However, we do note that there is no explanation for the discrepancies. Neither is there justification for the use of the graphemes that are proposed in the harmonized orthography. These are areas that our study addresses.

Mwangi et al. (2013) study how the orthography of Gĩkũyũ and Ekegusii can be harmonized. They identify intra- and inter-language discrepancies in the graphemic representation of both consonantal and vocalic phonemes and suggest graphemes that can be used instead. They further give advantages of harmonizing the two orthographies. This work was relevant because it discusses the Gĩkũyũ phonemic entry and also identifies some of the discrepancies that were of interest to us. We differed in that while their aim was harmonization of the two orthographies, ours was to account for the discrepancies in the orthography of Gĩkũyũ.

2.2. Theoretical Framework

2.2.1 Introduction

There are a number of phonological theories that this research could have chosen. Some of them are Generative Phonology (GP) (Chomsky and Halle, 1968), Metrical Phonology (Hayes, 1980), Autosegmental Phonology (Goldsmith, 1990) and Optimality Theory (Prince and Smolensky, 1993). However, the Correspondence Theory of the Optimality Theory appears best suited to investigate orthographic phenomena that, on the surface, look universal but, on close examination, are language specific.

2.2.2 The Optimality Theory

The Optimality Theory (OT) was originally proposed by Prince and Smolensky (1993) but has undergone subsequent developments. It was in reaction to the

inability of the then existing theories (SPE, Chomsky 1968; AP, Goldsmith, 1990, etc.) to account for phenomena that appeared universal yet did not hold in all languages.

The Optimality Theory works on the premise that each linguistic segment has an input and an output. The input in any language is unlimited. Prince and Smolensky (1993) present OT as a set of three interrelated modules: GEN, H-EVAL and Lexical Optimization. GEN and H-EVAL together comprise the grammar proper. Lexicon Optimization is not part of the grammar. It stores all the lexical forms of a language and provides the input specifications that are input to GEN (Kager, 1999). The formulations as set out by Prince and Smolensky for the GEN and H-EVAL modules are (p. 4):

GEN (In_k) $\rightarrow$ {Out_1, Out_2, …}

H- EVAL (Out_1, $1 \leq i \leq oo$ $\rightarrow$ Out_{real}

GEN is the operational component. For any given input (In_k), it generates a set of Candidate Output forms which may vary from the input in several ways. It applies its operations freely, optionally and sometimes, repeatedly. This is referred to as the freedom of analysis. Since GEN is universal within all languages, it anticipates all options that any language could transform the input into.

H-EVAL takes the Output Candidate forms from GEN and evaluates them against a set of constraints (CONS). It selects the most harmonic, most optimal as the

output of that language. The constraints are universal and violable. They are of two types: those that evaluate surface forms with respect to well-formedness (i.e. markedness), and those that deal with faithfulness. Markedness constraints are sensitive to the well-formedness of candidate form, banning disprefered or marked structures or requiring felicitous or unmarked properties. Faithfulness requires that the input (the underlying form of lexical entries) should be identical to the output (the surface form) of the respective input. Constraint interaction consists in the application of all the constraints, which are put into an order of dominance (the constraint hierarchy), to the inputs. Constraints have conflicting demands on a form. This is recognizable by making constraints violable though they can be violated only if this leads to the non-violation of a higher-ranking constraint. Their ranking varies between languages thus making languages different. This is exemplified in Table 2.1 below.

Table 2. 1: Relation of Dominance between Constraints

Input	Constraint 1	Constraint 2
Candidate A		*
Candidate B	*!	

(Source: Wiese 2010, p. 90)

Candidate A has not violated Constraint 1 and so is ranked higher. But if both candidates had violated Constraint 1, then Constraint 2 would be decisive in

determining the hierarchy if the higher ranked Constraint 1 cannot have an effect on the evaluation because both candidates violate Constraint 1 to the same extent. This is shown in Table 2.2 below.

Table 2. 2: The Decisive Role of Candidate 2

Input	Constraint 1	Constraint 2
☞Candidate 1	*	
Candidate 2	*	*!

(Source: Wiese 2010, p. 90)

Candidate 1 would be the preferred output.

Using the Optimality Theory, it is envisaged that there are five levels of representation in this study where the output is the grapheme while the underlying form is the phoneme.

In the first level, there is the Underlying Phoneme, the phoneme before it has undergone any phonological processes. This level corresponds to the phonological input (In_k) in OT. To give an illustration in Gĩkũyũ, the underlying phoneme could be the voiced labial plosive /b/.

In the second level, there is the Surface Phoneme. This is the level of the normal phonological output (Output1) of OT. It corresponds to phonemes as they are realized in given environments after undergoing phonological processes. From

our example, the phoneme in Gĩkũyũ is never realized without pre-nasalization and therefore will be realized as /^{m}b/.

The third level is an intermediate one where the phonemes, through some property-maintaining process, change into the orthography.

The fourth level is the Underlying Orthography. This is the writing system as designed by the designers of Gĩkũyũ orthography who have taken into consideration the principles of orthography design. The principles are the constraints in which an orthography is based (Baroni, 2016). The phoneme should correspond to the orthography. In our example, /^{m}b/ should appear graphemically as <mb>.

The last level is the Surface Orthography (Output 2). This is where the graphemes appear as presented by individual writers in the language, each with their own idiosyncrasies. From our illustration, depending on the ranking and, possibly violation of the principles in level four, the writers may present <mb>, <b>, <p>, <v> or <f>. These graphemes form the candidate set for our analysis.

These levels are shown in the Table 2.3 below.

Table 2. 3: Levels of Representation

LEVEL	CHARACTERISTICS
I Underlying phonemic	Phonological input
II Surface phoneme	Phonemes as realized by speakers
III Intermediate level	Phonemes change into graphemes
IV Underlying graphemic level	Orthography as designed
V Surface graphemic level	Graphemes as presented by authors

From the above schemata, it is apparent that the surface phoneme, after going through certain Constraints in level IV, should correspond to the designed orthography which authors should present (level V). The violation of the Constraints then is responsible for graphemic discrepancies. Constraints are universal but their ranking varies between languages. In this study, we identify and rank the constraints since 'it is the task of the linguist to find a constraint hierarchy which leads to the choice of the orthographic output which is correct in the language under consideration" (Wiese & Song, 2010; see also McCarthy, 2007, p. 8). To achieve the identification and ranking of the constraints, we have used the Correspondence Theory.

2.2.3 The Correspondence Theory

The Correspondence Theory is the notion that there is a correspondence relation ($\mathcal{R}$) between elements of each input and output level. The Theory is seen as the standard theory of faithfulness within Optimality Theory (van Oostendorp, 2004).

Correspondence is a relation between two structures, for example the base and reduplicant, or between the input and output. McCarthy and Prince (1995, p. 14) succinctly put it as:

> **Correspondence** Given two strings S1 and S2, correspondence is a relation $\mathcal{R}$ from the elements of S1 to those of S2. Elements $\alpha \in$ S1 and $\beta \in$ S2 are referred to as correspondents of one another when $\alpha \mathcal{R} \beta$. (p 23)

Correspondence is presented as a relation rather than a function to allow for one-to-many relations, such as diphthongization or coalescence.

The reduplicants or outputs are evaluated together with the base or input. Each candidate pair (S1, S2) comes from Gen equipped with a correspondence relation between S1 and S2 that expresses the relation between them. There is a correspondence relation for each (I,O) candidate-pair. EVAL then considers each candidate pair with its associated correspondence relations, assessing the completeness of correspondence in S1 or S2, the featural identity of correspondent elements in S1 and S2, and so on.

Constraints assess correspondence and identity of correspondent elements. For every correspondence relation, we have separate constraints. The separateness makes the constraints rankable. Correspondence constraints are given below:

1. The MAX Constraint Family

   Generally, these constraints ensure that every segment of S1 has a correspondent in S2. They include:

   MAX-BR where every segment of the base has a correspondent in the reduplicant which ensures that reduplication is total.

   MAX-IO where every segment of the input has a correspondent in the output. Thus, there is no phonological deletion and there is completeness of reduplicative copying.

2. The DEP Constraint Family

   These constraints ensure that every given segment of S1 has a correspondence in S2 so that S2 is ‘dependent on S1. The constraints are:

   DEP-BR in which case, every segment of the reduplicant has a correspondent in the base, and, DEP-IO which requires that every segment of the output has a correspondent in the input and so there is no epenthesis.

3. The IDENT (F) Constraint Family

   The constraints ensure that correspondent segments are identical in feature (F). The constraints are:

IDENT-BR (F) where the reduplicant correspondents of a base [F] are also [F].

IDENT-IO (F) the output correspondents of an input [(F] segment are also [(F]

When mismatches between corresponding elements in the different levels occur, they are constraint violations. Such violations are found but only when they are enforced by higher ranking constraints. Mismatches consist in omissions and/or additions of elements on one level, but not the other. A few examples in line with Wiese & Song's (2010) VISIBILITY and SIMPLICITY constraints are given below.

a) VISIBILITY constraints

1. Phonemes correspond to graphemes (for every phoneme, there is one grapheme).
2. Graphemes correspond to phonemes (for every grapheme, there is one phoneme).
3. Distinctive information corresponds to something in the orthography (phonemic vowel length is expressed orthographically)

b) SIMPLICITY constraints are given in the negative (Wiese, 2004) (indicated with an asterisk *):

1. *COMPLEX* every grapheme must have only one letter

2. *COMPLEX [DISCONNECTED] *letter shape must be continuous

In this study on the Gĩkũyũ language, we were, therefore, examining correspondence between phonological input and graphemic output.

2.2.4 Justification for use of the Optimality Theory

Many phonological studies on Gĩkũyũ have been done using theories based on Universal Grammar. These theories use rule ordering where the features under description are given in mathematical symbols to the effect that a process on one side of the equation will have a given output. Such approaches were effective in accounting for phenomena that were found in the underlying forms and realized in the surface forms. But as Prince and Smolensky (1993) observe, the rules are too powerful, unconstrained and redundant. Furthermore, they allow duplication and conspire in that different rules can have the same aim because they are not specific enough in what they try to achieve. We concur with Prince and Smolensky (1993) and feel that it is logical to avoid rules that are unnecessarily constraining. Multiple coexistent conflicting orthographic principles are better examined using OT rather than inviolable rules.

Phonological studies have also been carried out on Gĩkũyũ using representational/ non-linear theories such as Autosegmental and Metrical theories (Mwihaki, 1998; Kuria, 2005). OT is an offshoot of these theories and so should be able to analyze phonological – orthographical phenomena in a better way. We agree with Omachonu (2008, p. 95) who notes that ‘OT is adjudged as the most current,

efficient, viable and dependable because it is seen to have more explanatory force'.

None of the other theories is capable of showing that the constraints under study are universal but violable and that Gĩkũyũ re-ranks them to produce its own form (orthography). The Correspondence Theory of the Optimality Theory delves into the forms to expose the relationship between constraints on output structures and the operations that transform inputs into outputs. In this research, the underlying forms (Input) will be shown to have undergone some evaluation so as to be realized in certain forms in the surface structure (Output).

2.3 Chapter Summary

This chapter has presented the Literature Review and the Theoretical Framework. We have reviewed literature on Orthography Principles noting that these can be compressed into three criteria: Phonological Adequacy, Community Ownership, and Simplicity. Literature on Orthographical-Phonological Studies has revealed the interface between the two disciplines of orthography and phonology. We have reviewed literature on the Optimality Theory and orthography, and have observed that the theory is applicable. Lastly, we have reviewed literature about Orthography Studies on Kenyan Bantu Languages. In the Theoretical Framework, we have discussed the Correspondence Theory, a sub-theory of the Optimality Theory, and shown how it is used in this study. We now proceed to the next chapter, where we present our research methodology.

CHAPTER THREE

METHODOLOGY

3.0 Introduction

This chapter lays out the research design, the site of study, the sampling techniques and the sample size, the research instruments, and the data collection. It also includes the data presentation and analysis, as well as ethical considerations.

3.1 Research Design

This research used a qualitative design approach. Such a design "is concerned with how the social world is interpreted, understood, experienced, produced or constituted" (Mason, 2002, p. 3). Words, rather than numbers, are generated within the context of their natural occurrence as data for analysis and interpretation. The design was best suited for this research which aimed at gaining understanding of the graphemic structure of Gĩkũyũ within the natural set-up of the language. Phonemes and graphemes, rather than numbers, were generated, described and their characteristics analysed and interpreted using the Correspondence Theory.

3.2 Site of Study

Both phonological and orthographic data for this study were elicited from native speakers of Gĩkũyũ in each of the dialects: Gĩcũgũ, Mathira, Ndia, Northern Gĩkũyũ, and Southern Gĩkũyũ. Gĩcũgũ and Ndia dialects are found in Kirinyaga

County, Mathira and Northern Gĩkũyũ are in Nyeri County, while Southern Gĩkũyũ is in both Murang'a and Kiambu Counties. The specific locations for data collection were Kianyaga in Kirinyaga East sub-county (Gĩcũgũ) and Kagumo in Kirinyaga sub-county (Ndia), East Mathira sub-county (Mathira), Tetu sub-county (Northern Gĩkũyũ), and Kikuyu sub-county (Southern Gĩkũyũ).

3.3 Sampling Techniques and Sample Size

The total number of respondents for this research was thirty (30) from the five dialects. From each dialect, there were six (6) speakers. These were three (3) males and three (3) females. Gender differentiation is important because of the contention that language use varies with gender (Labov, 1991). The sample of thirty provided enough and reliable data because, as Wardhaugh (2006 p. 160) notes, large 'samples tend to be redundant, bringing increasing data handling problems with diminishing analytic returns'. Each respondent had lived in his/her locality since childhood and was aged over 40 years. This ensured that respondents had only one dialect of Gĩkũyũ. The age was important since orthographies are developed with mature readers in mind (Schroeder, 2010, p. 65). The respondents had proficiency in reading and writing Gĩkũyũ. In lower primary school, they had received instruction through Gĩkũyũ and had also learnt the language as a subject. They also had post- secondary school education which ensured that they could read and understand English from which they were to translate words into Gĩkũyũ. Purposive sampling was done to identify the actual respondents. The snowballing technique was employed to make sure that they

were from areas where there was bound to be less dialect levelling, such as urban centres (Meyerhoff, 2006; Wardhaugh, 2006).

The information above is depicted in the table below.

Table 3. 1: Respondents by Gender and Dialect

Dialect	Males	Females
Gĩcũgũ	3	3
Mathira	3	3
Ndia	3	3
Northern Gĩkũyũ	3	3
Southern Gĩkũyũ	3	3
Total by Gender	**15**	**15**
Grand Total	**30**	

3.4 Research Instruments

This study utilized two research instruments: questionnaires and a tape recorder. The questionnaire was most appropriate for eliciting data on both readability and writability in Gĩkũyũ, as well as graphemic data, while a tape recorder was best suited for capturing phonetic/ phonemic data which was later transcribed by the researcher.

3.4.1 Questionnaires

A questionnaire was presented to each of the six (6) respondents (See Appendix 1). Section One of the questionnaire had seven items. The first three items were based on the Likert Scale while the other four items required the respondent to give their own responses as requested.

The questionnaire sought to elicit data on the ease of readability and writing in Gĩkũyũ, and the respondents' comfort in the use of the orthography. The responses provided data on the principles of adequacy, simplicity and community ownership.

3.4.2 Transcripts from Tape Recordings

The informants were given a reading task and asked to translate and write the trnslations of specific words from English into Gĩkũyũ (see § Two, Appendix 1). They were asked to read their translations and were tape recorded as they did so (see § Three, Appendix I). The researcher transcribed the recordings so as to show the correlation between the phonemes and the graphemes. This provided data on phonological adequacy and simplicity principles.

3.5 Data Collection Procedures

A text was presented to the informants (See § Two, Appendix 1). The text was made up of English carrier sentences (Coupe, 2014) with purposively selected words to be translated into Gĩkũyũ (see Appendix II). The translations showed how respondents represented the phonemes graphemically.

Gĩkũyũ has an open syllable structure and minimally consists of a vowel (V) or a consonant-vowel (CV) structure (Githiora, 2016). Purposive sampling ensured that the Gĩkũyũ translations had the target consonantal graphemes in word initial and in word medial positions. This produced 36 graphemes. When the sentences were served to the thirty (30) informants, there were a total of 1080 consonantal graphemic tokens (see Table 3.2 below). These constituted our consonantal Candidates for evaluation using OT.

To get data on vocalic graphemes, specific words in English carrier sentences were presented to the thirty (30) informants to be translated and be written in Gĩkũyũ. In line with the syllable structure of Gĩkũyũ, the target vocalic graphemes were in the word initial, word medial and word final positions. The seven (7) graphemes multiplied by the three word positions and written by the thirty informants produced six hundred and thirty (630) vocalic graphemic tokens. They then read their translations and were tape recorded as they did so.

In Gĩkũyũ, vowel length is contrastive (Githiora, 2003). Words in carrier sentences were presented to the subjects. They were asked to translate and write the translations down, with the expectation that the target phoneme would be represented with a single grapheme in the word initial position (as required in Gĩkũyũ orthography), and as double vocalic graphemes in the medial and final word positions after translation. The doubled graphemes produced 630 tokens.

The targeted vocalic data is exemplified in Table 3.2 below. The respondents were then required to read their translations as they were being recorded.

Table 3. 2: Target Data

		Consonantal Graphemes	Vocalic Graphemes	
			Single	Double
Number		18	7	7
Word Position	Initial	1	1	1
	Medial	1	1	1
	Final	0	1	1
Respondents		30	30	30
Graphemic Data		**1080**	**630**	**630**

There are twenty-four (24) possible vocalic graphemic combinations according to Armstrong (1967) (see § 1.1). A close examination of the language by the researcher reveals that there are some other ten possible combinations, which makes a total of thirty-four (34). Out of these combinations, only six (6) are possible in the word initial position, twenty eight (28) in the medial position, and thirty two (32) in the word final position. Words in English were presented to the respondents and their translations were expected to contain the graphemes in the respective word positions. The respondents translated and wrote down their

translations. In the initial position, the thirty respondents generated one hundred and eighty (180) tokens. In the word medial position, they produced eight hundred and forty (840) tokens, and in the word final position, nine hundred and sixty (960) tokens were generated. Thus, the total number of tokens generated for the vocalic graphemes in concatenation were one thousand, nine hundred and eighty (1980) tokens.

Table 3.3 below summarizes the data collection for concatenated vocalic graphemes. After translation, the respondents read their translations and were recorded.

Table 3. 3: Targeted Concatenated Vocalic Graphemes

Word Position	Number of Graphemes	Respondents	Number of Tokens
Initial	6	30	180
Medial	28	30	840
Final	32	30	960
Total Number of Tokens			1980

The respondents wrote the graphemes which provided the researcher with graphemic data. These graphemes constituted our Candidates for evaluation using the Correspondence Theory of OT.

After each respondent had done the translation, he/she was asked to read each Gĩkũyũ word that had been translated. They were tape recorded as they read and the recordings were later transcribed by the researcher. The readings elicited 1080 tokens of consonantal phonological data and 3240 tokens of vocalic phonological data. The transcriptions were used to determine the phonological output that the informants represented graphemically. These provided data on phonological adequacy.

3.6 Data Presentation and Analysis

The phonemic and graphemic inventory of Gĩkũyũ are presented in tables and in figures, and then discussed. Data on the ease of writing and reading in Gĩkũyũ, the graphemes that are viewed as confusing, and on the graphemes that respondents would want included or excluded from the graphemic inventory are also presented in tables and subsequently discussed.

Consonantal graphemes are grouped according to the phonemes they represent. Thus, there are the plosives, the nasals, the fricatives, the affricates and the approximants. Each group is presented in tables showing the individual grapheme against the different realizations; that is, the candidates generated by the GEN component of OT. A discussion follows each tabular presentation.

Tables are also used to show the vocalic graphemes against the different realizations. The tables are used separately for the short vowels, the long vowels

(usually indicated by a double vowel), and for the vowels in concatenation. Each table is followed by a discussion.

Each grapheme is subsequently discussed in light of the Principles of Orthography Development. The relevant OT constraints are derived from the principles. Their hierarchy is established and then they are used to analyse the graphemes. The analysis is done using tableaux pitting the candidates (the generated graphemes) against the constraints so as to determine the optimal candidate, which is the winner.

The data is then examined under underrepresentation and overrepresentation so as to expose the impact that these two have on the speakers/ writers of Gĩkũyũ.

3.7 Ethical Considerations

The researcher sought, and obtained authorization from the Graduate School of Kenyatta University (see Appendix III). This in turn helped him to obtain permission to carry out the field study, from the National Council for Science and Technology (see Appendix IV).

The purpose of the study was explained to the respondents in advance and they were assured that the information they would give would not be disclosed to third parties. This ensured their informed consent. To ensure their anonymity, no names or addresses were used in the questionnaires or in the final thesis.

In addition, all sources of data were duly acknowledged.

3.8 Chapter Summary

In this chapter, we have discussed the research methodology employed in this study. We have presented the site of study, the sampling techniques used and the sample size. We have also given the instuments used in the study, the ways in which data has been collected, presented and analysed. Lastly, we have shown the ethical considerations undertaken in the study. We proceed to the next chapter in which the findings are presented and the data is analysed.

CHAPTER FOUR

PRINCIPLES OF ORTHOGRAPHY, THE CONSTRAINTS AND THE GĨKŨYŨ GRAPHEMES

4.0 Introduction

In this chapter, we analyse the graphemes generated by our respondents against the criteria used in orthography and against the Optimality Theory constraints. First, we present the inventories of Gĩkũyũ phonemes and their graphemes. This is followed by a presentation of each of the phonemes and the graphemes that were generated by the study. Next, we discuss each of the orthography criterion and its principles, and derive OT constraints from the principles. We then weigh the phonemes and graphemes against the principles and also evaluate the graphemes against the constraints. Finally, we provide a summary of the discussion and present our conclusion.

4.1 Gĩkũyũ Consonant Phonemes and Graphemes

Gĩkũyũ has eighteen consonantal phonemes. Four of the phonemes are prenasalised and four are nasal. Four other phonemes are voiceless. Of the remaining six, three are voiced fricatives, two are approximants and the last one is a tap. The table below shows the phonemes. Where they appear in pairs, the voiceless counterpart is to the left.

Table 4. 1: Gĩkũyũ Consonant Phoneme Inventory

	bilabial	Dental	alveolar	Palatal	velar	glottal
Stop	^{m}b		t ^{n}d		k ^{ŋ}g	
Nasal	m		n	ɲ	ŋ	
Tap			ɾ			
Fricative	β	ð		ʃ	ɣ	h
Affricate				^{ɲ}ɟ		
Approximant	w			J		

Adopted from Kioko et al. (2012, p. 45)

The palatal affricate /^{ɲ}ɟ/ is presented by some linguists as /^{ɲ}dʒ/ (Kioko et al., 2012; Kuria, 2005; Mwangi, 2003).The alveolar tap/flap /ɾ/ is also presented in some literature as /r/ (Armstrong, 1967; Benson, 1964; Mwangi et al., 2013; Kioko et al., 2012). In this study, we adopt /ɟ/, the symbol for the voiced palatal affricate in the IPA chart, and add the prenasal segment, /^{ɲ}/, to it. Thus, our symbol for the affricate is /^{ɲ}ɟ/. As for the alveolar tap/ flap, we use /ɾ/, the symbol given in the IPA chart.

Each of the consonants has a corresponding graphemic representation. The table below captures the representations.

Table 4. 2: Gĩkũyũ Consonant Grapheme Inventory

	bilabial	Dental	alveolar	palatal	Velar	Glottal
Stop	mb		t nd		k ng	
Nasal	m		n	ny	ng'	
Tap			r			
Fricative	b	th		c	g	h
Affricate				nj		
Approximant	w			y		

Adopted for Kioko, et al. (2012, p. 46)

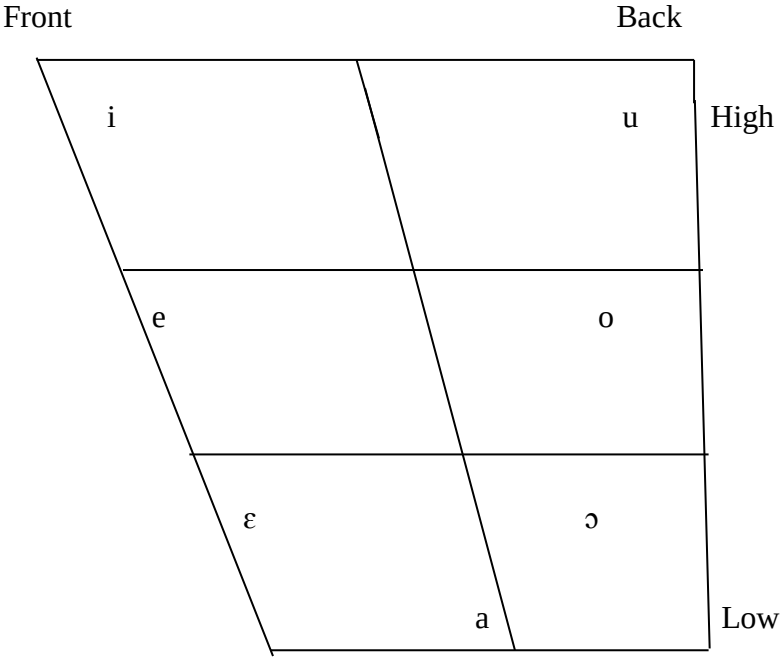
4.2 Gĩkũyũ Vocalic Phonemes and Graphemes

In this section, we first present the Gĩkũyũ short vowels and their graphemes followed by the long vowels and their graphemes too.

4.2.1 Gĩkũyũ Short vowels

There are seven short vocalic phonemes in Gĩkũyũ according to Armstrong (1967). These are /i/, /e/, /ɛ/, /a/, /ɔ/, /o/ and /u/. Using the criteria of muscular tension, lip spreading and height, she points out that Gĩkũyũ vowels lack crispness and precision (p. 4) and that they do not quite conform to Ladefoged's vowel trapezium (p. 8). Figure 4.1 below shows the vowels.

Figure 4. 1: Gĩkũyũ Vowels



Adopted from Kioko et al. (2012, p. 42)

Many linguists agree with this description by Armstrong, among them Mwangi, Njoroge & Mose (2013) and Kioko (2012). However, others such as Clements (1991) view the feature [height] as not adequate enough to describe Bantu vowel systems that have more than 3 height vowels. He argues that in a 4 - height vowel system like that of Gĩkũyũ, the height 3 vowels (/e/, /o/ and /ɛ/, /ɔ/) are sometimes described as [+high] and at other times as [-high]. Since [high] and [low] only describe three heights, a fourth height is not described (p. 35), the third height being [mid]. He proposes the use of an additional feature — [ATR].

[ATR] is a feature that uses the position of the tongue root to describe vowels. The tongue root can be advanced towards the front, [+ATR] or it can be retracted,

[-ATR]. Hulst (2014) observes that “advancing the tongue root inevitably leads to raising (or fronting) the tongue body and tongue root retraction has the opposite effect” (p. 14). ATR is contrastive (Kang and Kos, 2011). One contrast that appears is that [+ATR] vowels have lower F1 than their [-ATR] counterparts. This means that there is a contrast in the height of the vowels with [+ATR] vowels being higher. Another contrast is that, in their phonation, [+ATR] vowels are breathy as compared to [-ATR] vowels which are creaky (p. 1).

Peng (2016, p. 1) notes that many African languages have vowel contrasts characterized by [ATR]. He identifies the seven - vowel system as the most commonly occurring system and presents the system as shown in Figure 4.2 below.

Figure 4. 2: The Seven Vowel system

[+ATR]	/i/	/u/
[+ATR]	/e/	/o/
[-ATR]	/ɛ/	/ɔ/
[-ATR]	/a/	

Adopted from Casali (2016)

The description offered by Casali (2016) conforms closely to the vowels of Gĩkũyũ. Githiora (2016) notes that the two [+ATR] vowels /e/ and /o/ are more centralized than the high [+ATR] vowels /i/ and /u/. On his part, Peng (2000)

observes that in Gĩkũyũ, there is only one low vowel, /a/, and that the feature [+low] is sufficient to identify it. He additionally observes that mid vowels [+ATR] do not mix with mid vowels [-ATR].

Based on the above observations, in this research, we have adopted Armstrong's (1967) typology of 'front' and 'back' and also adopted Casali's (2016) typology of [±ATR] for purposes of identifying and discussing the vowels.We refer to the low [-ATR] vowel /a/ as the [Low] vowel in line with Peng's (2000) description. We therefore, propose the description shown in Figure 4.3 below.

Figure 4. 3: The Gĩkũyũ Short Vowels

	Front		Back	
[+ATR]	/i/		/u/	High
[+ATR]	/e/		/o/	Mid
[-ATR]		/ɛ/	/ɔ/	Mid
[-ATR]		/a/		[Low]

There are seven vocalic graphemes corresponding to the seven vowels. They are presented in Figure 4.4, each to the right of the phoneme it represents.

Figure 4. 4: Gĩkũyũ Vocalic Graphemic Inventory

[+ATR]	/i/ <i>	/u/ <u>	High
[+ATR]	/e/ <ĩ>	/o/ <ũ>	Mid
[-ATR]	/ɛ/ <e>	/ɔ/ <o>	Mid
[-ATR]	/a/ <a>		[Low]

4.2.2 Gĩkũyũ Long Vowels

Vowel length in Gĩkũyũ is significant. It is used for lexical differentiation. For example, *tara* 'count' and *taara* 'advise' are semantically differentiated by vocalic length. Phonemically, length is indicated through the phonetic notation (:). The long vocalic phonemes are shown in the figure below.

Figure 4. 5: Gĩkũyũ Long Vowels

[+ATR]	/i:/	/u:/	High
[+ATR]	/e:/	/o:/	Mid
[-ATR]	/ɛ:/	/ɔ:/	Low
[-ATR]	/a:/		[Low]

Orthographically, vowel length is represented by the doubling of each vocalic grapheme except in the word initial position. For example, in the word *anake* 'young men', the initial vocalic sound is loud and the word is pronounced as [aanakɛ]. This is also the convention employed in the development of new

orthographies (Stegen, 2005). Figure 4.6 below shows the long vocalic graphemes.

Figure 4. 6: Gĩkũyũ Long Vocalic Graphemes

[+ATR]	<ii>		<uu>	High
[+ATR]	<ĩĩ>		<ũũ>	Mid
[-ATR]	<ee>	<oo>		Mid
[-ATR]		<aa>		[Low]

4.3 The Data

To collect data on the Gĩkũyũ phonemes and graphemes, 30 respondents were targeted. They were educated mature three males and three females, from each of the Gĩkũyũ dialects of Southern, Northern, Mathira, Ndia and Gĩcũgũ (see § 3.4 above).

Through a questionnaire, the respondents were required to indicate the ease with which they write Gĩkũyũ. The responses ‘I strongly agree’ and ‘I agree’ were taken to mean ‘easy’, while ‘I disagree’ and ‘I strongly disagree’ were taken to mean ‘hard’. Data on the respondents who indicated that it is easy to write in Gĩkũyũ is presented in Table 4.3 below.

Table 4. 3: Gĩkũyũ is Easy to Write

Sex	Gĩcũgũ	Ndia	Mathira	Northern	Southern	Total
Male	3	2	2	2	1	10
Female	3	3	3	2	2	13

From the table above, all the female respondents in Gĩcũgũ, Ndia and Mathira indicated that writing in Gĩkũyũ was easy for them. In the Northern dialect, two males and two females indicated it was easy, while in the Southern dialect, only one male and two female respondents found it easy. The number of females affirming ease of writing is thirteen (13) which makes 86.7% of the total female respondents. For males, Gĩcũgũ and Mathira had the same number of respondents (three and two, respectively) affirming ease of writing. The male respondents reporting the same were fewer than the females in Ndia, Mathira and Southern dialects (two, two and one, respectively). The males who indicated ease of writing in Gĩkũyũ were ten (10). This is 66.7% of the total male respondents. Though gender was not a variable in this study, from the data, we surmise that females find writing in Gĩkũyũ easier than their male counterparts.

Another question required the subjects to indicate the ease with which they read Gĩkũyũ. The responses ‘I strongly agree’ and ‘I agree’ were interpreted as ‘easy’,

while ‘I disagree’ and ‘I strongly disagree’ were taken to mean ‘hard’. In Table 4.4, we present data from the respondents who said that reading Gĩkũyũ is easy.

Table 4. 4: Reading Gĩkũyũ is Easy

Sex	Gĩcũgũ	Ndia	Mathira	Northern	Southern	Total
Male	1	1	1	2	2	7
Female	2	3	2	1	2	10

We observe from the table above that only a single male respondent in each of the dialects of Gĩcũgũ, Ndia and Mathira found reading Gĩkũyũ easy, while two males in the Northern and in the Southern dialects reported the same. The total number of male respondents answering in the affirmative was seven (7), giving 47%. The female respondents who found reading Gĩkũyũ easy were two (2) each in Gĩcũgũ, Mathira and Southern, three (3) in Ndia and one (1) in Northern. This gave a total of ten (10) female respondents which translates to sixty seven percent (67%).

The subjects were also asked to indicate if reading and writing in English is easier than doing so in Gĩkũyũ. The responses ‘I strongly agree’ and ‘I agree’ were interpreted as ‘easier’, while ‘I disagree’ and ‘I strongly disagree’ were interpreted as ‘more difficult’. The responses of those answering ‘easier’ are presented in Table 4.5.

Table 4. 5: Reading and Writing in English Easier than in Gĩkũyũ

Sex	Gĩcũgũ	Ndia	Mathira	Northern	Southern	Total
Male	2	2	1	3	3	11
Female	3	3	2	2	1	11

As seen from the table above, all the male respondents in the Northern and the Southern dialects indicated that they preferred reading and writing in English to doing so in Gĩkũyũ. Two males in Gĩcũgũ and in Ndia indicated the same while in Mathira, there was only one male. This gave a total of eleven (11) male respondents and a percentage of seventy-three (73%).

All the female respondents in Gĩcũgũ and Ndia said that reading and writing in English is easier than doing so in Gĩkũyũ. In Mathira and Northern, two respondents in each reported the same, as did one in Southern. This was a total of eleven (11) respondents which is also seventy-three (73%) percent.

While in Gĩcũgũ, Ndia and Mathira there were more female respondents who were more comfortable reading and writing in English than in Gĩkũyũ (8 females against 5 males), the converse was true in the Northern and Southern dialects (3 females against 6 males).

A second task required the respondents to identify consonantal graphemes that they found most confusing, and to provide reasons as to why they were confusing.

The graphemes they identified were <mb>, <b>, <nd>, <ng>, <ng'>, <nj>, <ny>, <th>, <h>, <t> and <c>. Table 4.6 below shows the identified graphemes against the number, dialect and sex of respondents who identified them.

Table 4. 6: Confusing Consonantal Graphemes

Grapheme	Gĩcũgũ		Ndia		Mathira		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
<mb>	-	2	-	-	1	1	2	-	-	-	**3**	**3**
<b>	-	2	-	-	1	1	1				**2**	**3**
<nd>	-	-	-	-	1	-	-	-	-	-	**1**	**0**
<ng>	1	-	1	1	1	2	2	1	1	1	**6**	**5**
<ng'>	2	-	1	2	1	2	-	2	3	1	**7**	**7**
<nj>	-	-	-	-	-	1	-	1	1	-	**2**	**1**
<ny>	1	-	-	-	-	-	-	-	-	-	**1**	**0**
<th>	-	1	-	-	-	-	1	1	-	-	**1**	**2**
<h>	1	-	-	1	-	-	1	-	-	-	**2**	**1**
<t>	-	1	-	-	-	-	-	-	-	-	**0**	**1**
<c>	-	-	-	-	-	-	1	-	-	-	**1**	**0**

Legend: **M** = males **F** = females

Graphemes <mb> was identified as a confusing grapheme by six respondents made up of an equal number of males and females. This is 20% of all the respondents. In Mathira, there was one male and one female identifying the grapheme. Two females from Gĩcũgũ and two males from Northern identified the grapheme. No respondents from Ndia and Southern identified the grapheme. The respondents closely tied this grapheme to <b>.

Grapheme <b> was identified by five respondents, two females from Gĩcũgũ, a male and a female from Mathira, and one male from Northern. This was 13.3% of the total male respondents and 20% of the females. As in the case with <mb>, the grapheme was not identified by respondents from Ndia and Southern. The respondents from those three dialects who identified the two graphemes said that <b> and <mb> represented the same sound to them. They said that <b> is confusing because it is too removed from the sound [f] that they claimed it represents.

Grapheme <nd> was identified by one male respondent from Mathira. This was 6.7% of the male respondents. He did not give any reason why he found the grapheme confusing. No female respondents identified the grapheme.

Eleven respondents identified <ng> and <ng'> as confusing graphemes. These were respondents from all the dialects and cut across both sexes. Only the female respondents from Gĩcũgũ did not find the graphemes confusing. The reason given for the confusion between the two graphemes is that they are similar except for

the apostrophe. This made it difficult for the speakers to decide when to use either <ng> or <ng'>. The two graphemes are also confusing to them because in English, <ng> represents the sound [ŋ], the same sound that is represented by <ng'> in Gĩkũyũ, yet the two graphemes are too close in appearance. In addition, <ng'> does not exist in English.

Grapheme <nj> was identified by three respondents, a female from Mathira and another from Northern, and one male from Southern. They did not give any reason for identifying the grapheme.

One male respondent from Gĩcũgũ identified <ny> as confusing. He did not give any reason as to why he found it confusing.

Grapheme <th> was identified as confusing by three respondents. These were a female from Gĩcũgũ and another one from Northern in addition to a male from Ndia. One female respondent from Gĩcũgũ identified <t> as confusing. Grapheme <h> was identified as confusing by one female respondent from Ndia, and two males, one from Gĩcũgũ and the other from Northern. The respondents who identified graphemes <t> and <th> said that the two represent almost the same sound and they were unsure of when to use them. They also stated that the 'h' in <th> confused them since it appeared like a separate grapheme. The respondent who identified <h> said that it was confusing because she did not know exactly when to use it. Grapheme <c> was found confusing by one male respondent from Northern. He said the confusion arose because he was unsure when to use it.

A third task required the respondents to identify vocalic graphemes that they found difficult in reading. They identified <e>, <i>, <ĩ >, <o>, <u> and <ũ>, virtually all the vocalic graphemes except <a>. In Table 4.7 below, we present these graphemes against the number, sex and dialect of the respondents who identified them.

Table 4. 7: Confusing Vocalic Graphemes

Graphemes	Gĩcũgũ		Ndia		Mathira		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
<e>	1	-	-	-	2	1	-	1	-	-	**3**	**2**
<i>	-	-	-	-	1	-	1	1	1	-	**2**	**2**
<ĩ>	-	-	1	1	2	1	3	1	2	1	**8**	**4**
<o>	1	-	1	-	1	-	-	-	-	-	**3**	**0**
<u>	-	-	1	-	1	1	1	-	1	-	**4**	**1**
<ũ>	-	-	1	1	2	1	3	1	3	1	**9**	**4**

Legend: M = males F = females

The male respondents who identified the confusing graphemes were more than the females except in the case of <i> where there was a tie. From Gĩcũgũ, females did not identify any grapheme as confusing. From Ndia and Southern, a single female in each dialect identified only one and the same grapheme, that is, <ĩ>. Grapheme <ũ> was identified by nine (9) males which is 60%, and four (4) females which is 26.7%. This is closely followed by <ĩ> which was identified by

eight (8) males, giving 53.3%, and four (4) females at 26.7%. The implication from the above analysis is that females are more comfortable with the vocalic graphemes than the males.

Mathira, Northern and Southern had a higher number of male respondents identifying confusing graphemes (nine, eight and seven, respectively) as compared to Ndia and Gĩcũgũ (four and two, respectively). The same was also true of the females (four each from Mathira and Northern and two from Southern, respectively against two from Ndia and none from Gĩcũgũ).

The thirteen respondents who identified grapheme <ũ> said that it caused confusion with grapheme <u> and that it was not easy to differentiate the two. They added that they are not used to grapheme <ũ> because it is not found in English. The twelve respondents who identified grapheme <ĩ> said it was also not easy to distinguish it from <i>, and that it sounded the same as <i>. The five respondents who identified <e> said that it was confusing because it is pronounced differently in English.

Respondents were also required to indicate the graphemes they thought should be excluded from the Gĩkũyũ orthography and to provide reasons why they thought so. The graphemes that were identified were <mb>, <nd>, <ng>, <ng'>, <nj>, <th>, <y>, <h>, <ĩ> and <ũ>. Also identified were graphemes <j>, <l>, <p>, <s>, <q>, <v>, <x> and <z> though they are not in the graphemic inventory of

Gĩkũyũ. We present the responses in Table 4.8 below. The grapheme is shown against the number, sex and dialect of respondents.

Table 4. 8: Graphemes to be Excluded from the Inventory

Graphemes	Gĩcũgũ		Ndia		Mathira		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
<mb>	-	-	-	-	1	-	-	-	-	-	**1**	**0**
<nd>	-	-	-	-	1	-	-	-	-	-	**1**	**0**
<ng>	-	-	-	-	-	-	1	1	-	-	**1**	**1**
<nj>	-	-	-	-	-	-	1	1	-	-	**1**	**1**
<th>	1	-	-	-	-	-	1	-	-	-	**2**	**0**
<h>	-	-	-	-	-	-	1	-	-	-	**1**	**0**
<j>	-	-	-	-	-	-	-	-	1	-	**1**	**0**
<k>	-	-	-	1	-	-	-	-	-	-	**0**	**1**
<l>	1	-	-	1	-	-	-	-	-	-	**1**	**1**
<p>	1	-	-	1	-	-	-	-	-	-	**1**	**1**
<q>	1	1	-	1	-	-	-	-	1	-	**2**	**2**
<s>	2	1	-	1	-	-	-	-	1	-	**3**	**2**
<v>	-	1	-	1	-	-	-	-	-	-	**0**	**2**
<x>	1	-	-	1	-	-	-	-	-	-	**1**	**1**
<y>	-	-	-	-	1	-	-	-	-	1	**1**	**1**
<z>	3	1	-	1	-	1	-	-	1	-	**4**	**3**
<ĩ>	-	-	-	-	-	-	1	-	2	1	**3**	**1**
<ũ>	-	-	-	-	-	-	1	-	2	1	**3**	**1**

Legend: **M** = males **F** = females

The respondents who identified <mb>, <nd>, <ng> and <nj> suggested that the graphemes should not have the ‘n’ segment. The two who identified <th> said that it was too difficult to pronounce, the same reason given by the one respondent who identified <h>. About grapheme <y>, the respondents said that there were few words in which the grapheme is used. We interpreted this to mean that to the respondents, the grapheme has a low functional load in Gĩkũyũ. The graphemes <j>, <l>, <p>, <s>, <q>, <v>, <x> and <z> are not in the Gĩkũyũ inventory yet they were identified for exclusion in twenty - five (25) responses. We understood this to mean that the respondents had interacted with the graphemes in literature purportedly written in Gĩkũyũ and were unaware that the graphemes do not form part of the graphemic inventory. This might also explain the high percentage of respondents who expressed ease in writing Gĩkũyũ. On graphemes <ĩ> and <ũ> that were identified in eight responses, the respondents said that they caused confusion in both reading and writing. They added that the two graphemes are not found in the vowel system of Bantu languages that they are familiar with, such as Kiswahili.

When the respondents were asked to suggest the graphemes they would want included in the graphemic inventory, they gave <f>, <j>, <sh>, <s> and <b>. Table 4.9 below shows the responses to this question.

Table 4. 9: Graphemes to be Included in the Graphemic Inventory

Graphemes	Gĩcũgũ		Ndia		Mathira		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
<f>	-	-	-	-	-	1	-	-	-	-	**0**	**1**
<j>	-	-	-	-	1	-	-	-	-	-	**1**	**0**
<sh>	-	-	-	-	1	-	-	-	1	-	**2**	**0**
<s>	-	-	-	-	1	-	-	-	1	-	**2**	**0**
<b>	-	-	-	-	-	-	1	-	-	-	**1**	**0**

Legend: **M** = males **F** = females

We note that there were no suggestions from any Gĩcũgũ and Ndia respondents nor from the female respondents from Northern and Southern. In addition, only one female respondent from Mathira suggested <f>. The respondents who made suggestions said that those graphemes would actually represent the sounds that they produced. The one who identified <b> suggested that it should be used in place of <mb>.

The findings from the above three tasks are discussed in detail in Chapter 5.

In a separate task, the respondents were presented with some carrier sentences in English that contained words and phrases with the target phonemes and graphemes. They were asked to read the sentences then translate and write the

target words in Gĩkũyũ. This provided the graphemic data. They then read the words they had written and were taped as they did so. This provided us with the phonemic data. We did note that the subjects who had said that reading or writing in Gĩkũyũ is easy did not actually do so with the ease they had indicated.

Owing to the fact that one criterion relied on in the identification of the respondents was their ability to read and write in Gĩkũyũ, we expected to collect only a small sample of graphemes and phonemes that did not correspond to the graphemic and phonemic inventories. Consequently, the number of occurrence of the variant/ deviant phonemes or graphemes was inconsequential, but as Anthony (2013, p. 146) notes, "the value of a corpus is clearly dependent not on its size, but on what kind of information we can extract from it". The data we collected was, therefore, adequate for our study.

In the following sub-sections, we present the phonemic and graphemic data that was generated in the study. We shall first present the consonantal data then follow it up with the vocalic data.

4.3.1 The Plosives

Gĩkũyũ has two voiceless plosives, the alveolar /t/ and the velar /k/. Their graphemes are <t> and <k> respectively. The language has three prenasalised plosives which are the bilabial /^{m}b/, the alveolar /^{n}d/, and the velar /^{ŋ}g/. Their graphemes are <mb>, <nd> and <ng>, respectively.

The respondents were given words that contained the target phonemes in the initial and medial positions and asked to translate and write the words, thereby generating the graphemes corresponding to the phonemes.

The data below shows the graphemes that the respondents generated in the carrier words and their occurrence in the initial and medial word positions.

(1)

	Phoneme	Grapheme	Initial	Gloss	Medial	Gloss
a)	/^{m}b/	<mb>	**mb**athi	‘bus’	ng’o**mb**e	‘cow’
		<b>	**b**athi	‘bus’	-	
b)	/t/	<t>	**t**ũhe	‘give us’	ri**t**a	‘litre’
c)	/^{n}d/	<nd>	**nd**uka	‘shop’	mũgũ**nd**a‘	garden’
		<d>	**d**uka	‘shop’	-	
d)	/k/	<k>	**k**ĩrooto	‘dream’	n**y**o**k**a	‘snake’
e)	/^{ŋ}g/	<ng>	**ng**ai	‘God’	u**ng**umania	‘corruption’
		<g>	**g**ai	‘God’	-	

From the responses above, each of the voiceless plosives /t/ and /k/ has a single graphemic representation, <t> and <k>, respectively. No discrepancies with the designed orthography occur. In contrast, each of the prenasalised voiced plosives is represented with more than one grapheme in the word initial position, but a single grapheme in the medial position. Phoneme /^{m}b/ is represented with <mb> and <b> in the word initial position but with <mb> only in the medial position;

phoneme /^{n}d/ is represented with <nd> and <d> in the initial position but only <nd> in the medial position; and /^{ŋ}g/ is represented with <ng> and <g> in the initial position but with only <ng> in the medial position. This representation occasions discrepancies in the word initial position.

In Table 4.10 below, we present the graphemes that differed from the target graphemes. The target graphemes are <mb>, <nd> and <ng> while the variant graphemes were <b>, <d> and <g>, respectively. We indicate the number, sex and dialect of the respondents who gave the graphemes.

Table 4. 10: Respondents by dialect and sex against deviant plosive graphemes

Graphemes	Gĩcũgũ		Ndia		Mathira		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
<b>	0	2	1	0	0	0	0	1	2	0	**3**	**3**
<d>	0	0	1	1	0	1	0	1	2	0	**3**	**3**
<g>	0	0	1	0	0	0	0	0	0	0	**1**	**0**

Legend: **M** = males **F** = females

From the Table above, males in Gĩcũgũ, Mathira and Northern generated no discrepancies with the graphemes for plosives that are in the inventory. The female respondents in Southern also generated no discrepancies. Grapheme <ng> had the least discrepancies with only a single male respondent from Ndia. Mathira

respondents generated the least discrepancies with only one female producing <d> for <nd>.

In the reading task, the respondents elicited a single phone (surface realization) for each voiceless plosive. Phoneme /t/ was realised as [t] and /k/ as [k]. On the other hand, more than one surface realization was generated for the prenasalised plosives. The phoneme /^{m}b/ was produced as [^{m}b] and [b], /^{n}d/ as [^{n}d] and [d], and /^{ŋ}g/ as [^{ŋ}g] and [g]. Thus, each phoneme has two allophones. The prenasalised phones were most prevalent in the word medial positions while the non-prenasalised ones were most prominent in the word initial position. In Table 4.11 below, we give the phoneme and the differing phone(s), and the number, dialect and sex of respondents who produced the phone(s).

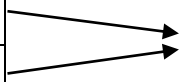
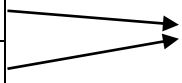
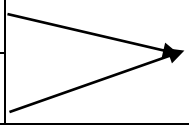
Table 4. 11: Differing Phones for Prenasalised Phonemes produced by Respondents

Phoneme / Phone	Gĩcũgũ		Ndia		Mathira		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
/^{m}b/ [b]	1	-	1	-	1	-	-	-	-	-	**3**	**0**
/^{n}d/ [d]	-	-	-	-	2	1	-	-	-	1	**2**	**2**
/^{ŋ}g/ [g]	1	1	-	-	1	1	-	-	-	-	**2**	**2**

Legend: **M** = males **F** = females

Table 4.12 below captures the information about the plosives. In the table (and subsequent ones in this sub-section that describe the phonetic realization of phonemes), the phoneme under study is given in the first column. The second column shows how that phoneme is realized on the surface (the phone or phones) by the respondents. The arrows in the third column represent the process where the phone(s) transit and converge into the specific grapheme which represents the phoneme (as it is given by the designers of Gĩkũyũ orthography). That specific grapheme is given in the fourth column. The last column contains the grapheme or graphemes as they were generated by the respondents (the surface graphemic representation). We view the specific grapheme as basic and the surface graphemes as allographs of the specific grapheme.

Table 4. 12: The Gĩkũyũ Plosives

phoneme	surface realization	Transition	grapheme	surface graphemic representations
/^{m}b/	[^{m}b]		<mb>	<mb>
	[b]			<b>
/t/	[t]		<t>	<t>
/^{n}d/	[^{n}d]		<nd>	<nd>
	[d]			<d>
/k/	[k]		<k>	<k>
/^{ŋ}g/	[^{ŋ}g]		<ng>	<ng>
	[g]			<g>

From the foregoing, it is apparent that discrepancies occur with the prenasalised plosives and not the voiceless plosives. The two sets of plosives differ firstly, in voice and secondly, in the nasal segment. When this nasal segment is not represented, and that is in the word initial position, discrepancies in writing occur. Ngure (2005) notes that in the word initial position, the nasal segment occurs underlyingly, while in the middle position, it is inherent, meaning that it is embedded in the root of the word. In the subsequent paragraphs, we discuss this underlying nasal segment since it is obviously significant in the graphemic representation. We shall refer to the nasal as /N/ in line with other linguists such as Peng (2003) and Choti (2015).

As already noted, in Gĩkũyũ, all the voiced plosives are preceded by the nasal segment /N/ (Peng, 2003; Jeon et al., 2015). In the word initial position, the underlying phoneme for the voiced bilabial plosive /^{m}b/ is the bilabial voiced fricative /β/. For the voiced alveolar plosive /^{n}d/, the underlying phoneme is the voiceless plosive /t/, while for the voiced velar plosive /^{ŋ}g/ it is the velar fricative /ɣ/ (Peng, 2003; Ngure, 2005; Farris-Trimble, 2008). The phonotactics of Gĩkũyũ does not allow a nasal segment to be followed by a voiced continuant, a voiceless stop or a fricative. Thus, we may not have /N+β/, /N+t/, /N+ ɣ/ or even /N+ʃ/. Consequently, several repair processes take place.

The voiced bilabial fricative /β/ undergoes hardening to become the bilabial plosive /b/. Hardening is the process in which a voiceless obstruent or a fricative

coming after a nasal is realised as a post-nasal stop (Peng, 2003, pp. 203-5). To minimize the articulatory effort, the nasal segment /N/ assimilates to /m/ which shares the same place of articulation with /β/. Massamba (1996) refers to this place assimilation as homorganic nasal assimilation (p. 93). The hardening of /β/ and the place assimilation of /N/ to /m/ results in the bilabial plosive /^{m}b/ physically realised as [^{m}b].

The voiceless alveolar plosive /t/ undergoes voicing and is realised as /d/. The nasal segment /N/ preceding it assimilates to the alveolar place of articulation so as to minimize the articulatory effort and is realised as /n/. The phoneme thus produced is /^{n}d/ and the phone [^{n}d].

The voiced velar fricative /ɣ/ undergoes hardening and is realised as [g]. To minimize the articulatory effort, the preceding nasal segment assimilates to the velar place of articulation and is realized as /ŋ/. The phoneme which is produced is /^{ŋ}g/ whose physical manifestation is [^{ŋ}g].

Armstrong (1967) notes that the nasal segment is weak in the word initial position and that in some cases, it is wholly absent. Ngure (2005) also observes the loss of this segment in Gĩkũyũ which he attributes to age, level of education and geographical factors. Other researchers do note the same too (Jeon et al., 2015 p. 3). However, Armstrong (1967) proposes that the segment should be represented graphemically. In this thesis, we agree with the researchers that the nasal segment,

whether present or absent in the initial or medial position, should be represented orthographically as it enhances homogeneity and lessens spelling errors.

4.3.2 The Gĩkũyũ Nasals

There are four Gĩkũyũ nasals, /m/, /n/, /ɲ/ and /ŋ/ which are graphemically represented as <m>, <n>, <ny>, and <ng'>, respectively. To generate graphemic data, subjects were given words to translate and write in Gĩkũyũ. All the respondents produced the same graphemes for /m/ and /n/, that is, <m> and <n> which are in the graphemic inventory of Gĩkũyũ. However, the graphemes for /ɲ/ and /ŋ/ had variants.

The target phonemes were in the word initial and medial positions. The results are presented below.

(2)

Phoneme	Grapheme	Initial	Gloss	Medial	Gloss
a) /m/	< m>	**m**ũgwanja	'seven'	kũ**m**eria	'swallow'
b) /n/	< n>	**n**ĩũi	'do you know'	a**n**ake	'young men'
d) /ɲ/	< ny>	**ny**au	'cat'	i**ny**a	'four'
	< nj>	**nj**au	'cat'	i**nj**a	'four'
c) /ŋ/	<ng'>	**ng'**ombe	'cow'	njo**ng'**i	'ugly'
	<ng>	**ng**ombe	'cow'	njo**ng**i	'ugly'

As seen in the data above, the bilabial nasal /m/ and the alveolar nasal /n/ are each represented by only one grapheme, <m> and <n>, respectively. There are no

discrepancies with the graphemic inventory. However, the palatal nasal /ɲ/ and the velar nasal are each represented with more than one grapheme in both word positions. This creates discrepancies.

In Table 4.13 below, we present the number of respondents, their sex and dialect against the deviant graphemes that they generated.

Table 4. 13: Respondents by Sex and Dialect against Deviant Nasal Graphemes

Graphemes	Gĩcũgũ		Ndia		Mathĩra		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
<nj>	0	1	1	1	0	1	0	0	0	0	**1**	**3**
<ng>	1	1	0	0	0	0	2	0	2	1	**5**	**2**

Legend: **M** = males **F** = females

The grapheme <nj> is presented by a single female respondent from each of the Gĩcũgũ, Ndia and Mathĩra dialects. One male from Ndia also presents <nj>. Northern and Southern dialects do not produce grapheme <nj>. Overall, males generate less variants in this grapheme. Ndia generates most discrepancies. For the velar nasal /ŋ/, the grapheme <ng> is presented by a male respondent from Gĩcũgũ, two males from Northern and two other males from Southern. In Ndia

and Mathĩra, there are no discrepancies. Overall, males produce more discrepancies than females.

When the respondents read their translations, they produced one surface (phonetic) realization for each phoneme except for /ɲ/. This phoneme was presented as [ⁿɟ] in the word initial position by one female respondent from Gĩcũgũ. The information on the phonemes and the phones is shown in Table 4.14 below.

Table 4. 14: The Gĩkũyũ Nasals

Phoneme	Surface representation	transition	Graphemic representation	Surface Graphemic representations
/m/	[m]	→	<m>	<m>
/n/	[n]	→	<n>	<n>
/ɲ/	[ɲ] [ⁿɟ]	→	<ny>	<ny> <nj>
/ŋ/	[ŋ]	→	<ng'>	<ng'> <ng>

As can be observed in the table, there are no discrepancies between the surface realizations of /m/ and /n/ and their respective graphemes. Discrepancies only

occur between /ɲ/ and /ŋ/ on one hand and their respective graphemes on the other. Each had two allographs to it.

4.3.3 The Palatal Affricate /ᵑɲ̟/, <nj>

The palatal affricate /ᵑɲ̟/ is represented graphemically by <nj>. The respondents generated the data below in word initial and medial positions.

(3)

	Phoneme	Grapheme	Initial	Gloss	Medial	Gloss
a)	/ᵑɲ̟/	<nj>	**nj**anji	'judge'	mũgwa**nj**a	'seven'
		<j>	**j**anji	'judge'	mũgwa**j**a	'seven'

The data shows that /ᵑɲ̟/ had two representations, <nj> and <j> in both word positions. This creates discrepancies in the orthography. In the following table, we show the deviant grapheme against the number, dialect and sex of the respondents who generated it.

Table 4. 15: Respondents Producing Variant Grapheme <j>

Grapheme	Gĩcũgũ		Ndia		Mathĩra		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
<j>	1	1	1	0	0	1	1	2	0	0	**3**	**4**

Legend: **M** = males **F** = females

The grapheme was presented by respondents from all the dialects except the Southern. There were also more females than males in its generation.

From the reading task, the palatal affricate had two surface realizations: [ⁿɟ] and [ɟ]. The second phone [ɟ] was given by respondents from all the dialects.

The data on the phoneme, its phones and graphemes, is given in Table 4.16 below.

Table 4. 16: The Palatal Affricate

Phoneme	Surface realization	Transition	Graphemic representation	Surface Graphemic representations
/ⁿɟ/	[ⁿɟ]	→	<nj>	<nj>
	[ɟ]	→		<j>

The voiced palatal affricate /ⁿɟ/ has its underlying form as the voiceless fricative /ʃ/ in the word initial position which makes it a derived phoneme (Ngure, 2005). The affricate is also preceded by the placeless nasal /N/. To conform to Gĩkũyũ phonotactics, the voiceless palatal fricative [ʃ] undergoes hardening to become the voiceless affricate [tʃ], and then it goes through voicing to become the voiced affricate [ɟ]. The nasal /N/ assimilates to the palate and so shares the same place of articulation with [ɟ] and is realized as [ɲ]. The phoneme thus realized is /ⁿɟ/ and its surface realization [ⁿɟ]. In the word medial position, the nasal segment is

embedded in the word (Ngure, 2005). Graphemically, the nasal segment should be represented with <n>. The non-representation of this segment leads to discrepancies in the orthography so that we get both <nj> and <j> as allographs.

4.3.4 The Fricatives

The Gĩkũyũ fricatives are the voiced bilabial /β/, the voiced interdental /ð/, the voiceless palatal /ʃ/, the voiced velar /ɣ/, and the voiceless glottal /h/. Their graphemic representations are <b>, <th>, <c>, <g> and <h>, respectively. This group of phonemes presents the highest number of variants in their graphemic representations as were seen when carrier words were given for translation and writing. The data below shows the responses.

(4)

	Phoneme	Grapheme	Initial	Gloss	Medial	Gloss
a)	/β/	<b>	**b**etũrũ	‘petrol’	ma**b**anga	‘machetes’
		<mb>	-		ma**mb**anga	‘machetes’
		<p>	**p**etũrũ	‘petrol’	ma**p**anga	‘machetes’
		<v>	**v**etũrũ	‘petrol’	ma**v**anga	‘machetes’
		<f>	-		ma**f**anga	‘machetes’
b)	/ð/	<th>	**th**ayũ	‘peace’	mba**th**i	‘bus’
		<nd>	**nd**ayũ	‘peace’	mba**nd**i	‘bus’
		<d>	-		mba**d**i	‘bus’
		<ny>	-		mba**ny**i	‘bus’
c)	/ʃ/	<c>	**c**iũria	‘questions’	ma**c**ungwa	‘oranges’

		<s>	**s**iuria	‘questions’	ma**s**ungwa	‘oranges’
		<ch>	**ch**iũria	‘questions’	ma**ch**ungwa	‘oranges’
d)	/ɣ/	<g>	**g**athenge	‘he-goat’	mbo**g**o	‘buffalo’
		<k>	**k**athenge	‘he-goat’	-	
e)	/h/	<h>	**h**uuhu	‘bats’	tũ**h**e	‘give us’
		<b>	**b**uubu	‘bats’	tũ**b**e	‘give us’
		<mb>	**mb**ubu	‘bats’	-	

Grapheme <b> which is in the graphemic inventory was presented in both word positions. Grapheme <mb> was only presented in the word medial position. Graphemes <p> and <v> were presented in both word positions while <f> was only given in word medial position.

For phoneme /ð/, in addition to <th>, graphemes <nd>, <d> and <ny> were also presented. Grapheme <nd> appeared in both medial and final word positions. Graphemes <d> and <ny> were only presented in the word medial position.

As for phoneme /ʃ/, graphemes <c>, <s> and <ch> were presented. The graphemes were presented in both word positions.

Phoneme /ɣ/ is represented with graphemes <g> and <k>. The deviant grapheme, <k> is presented only in the word initial position.

For phoneme /h/, the graphemes <h>, <b> and <mb> are presented. While grapheme <h> and the deviant <b> were presented in both word positions, deviant <mb> is presented in the word initial position only.

The following table shows the fricatives and the deviant graphemes that were generated. It also shows the number and sex of respondents who generated the graphemes and the dialects to which the respondents belong.

Table 4. 17: Respondents and Variant Fricative Graphemes

Phoneme/ Grapheme	Gĩcũgũ		Ndia		Mathĩra		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
/β/ <mb>	2	3	2	1	-	-	-	-	-	-	**4**	**4**
<p>	1	2	-	-	-	-	-	1	-	-	**1**	**3**
<v>	1	2	1	1	-	1	1	2	3	-	**6**	**6**
<f>	-	-	-	-	1	-	-	1	-	-	**1**	**1**
/ð/ <nd>	2	3	-	-	-	-	-	-	-	-	**2**	**3**
<d>	-	1	-	-	-	-	-	-	-	-	**0**	**1**
<ny>	1	-	-	-	-	-	-	-	-	-	**1**	**0**
/ʃ/ <s>	-	-	-	-	-	-	-	-	-	1	**0**	**1**
<ch>	-	2	1	-	3	3	3	-	2	-	**9**	**5**
/ɣ/ <k>	2	2	1	1	-	-	-	-	-	-	**3**	**3**
/h/ <b>	3	3	-	2	-	-	-	-	-	-	**3**	**5**
<mb>	2	3	2	1	-	-	-	-	-	-	**4**	**4**

Legend: **M** = males **F** = females

For phoneme /β/, graphemes <mb>, <p>, <v> and <f> were produced, creating discrepancies. Grapheme <mb> was generated by respondents from Gĩcũgũ and Ndia who included two males from each dialect, three females from Gĩcũgũ and one female from Ndia. Grapheme <p> was generated by one male and two females from Gĩcũgũ, and one female from Northern. Grapheme <v> was produced across all dialects. In Southern, all the males produced it but no females did. In Mathĩra, only one female generated it. Grapheme <f> was generated by a single male from Mathĩra and a female from Northern. All the deviant graphemes representing /ð/ were given by Gĩcũgũ respondents only. Grapheme <nd> for <th> was presented by all the females and two males. Grapheme <d> was produced by a female speaker while <ny> was presented by a male one.

As for phoneme /ʃ/, grapheme <s> was presented by one female speaker from the Southern dialect. Grapheme <ch> was generated by all the respondents from Mathĩra and all the males from Northern. There were no female speakers from Ndia, Northern and Southern who produced the grapheme and neither did any male from Gĩcũgũ.

Grapheme <k> for /ɣ/ is presented by speakers from Gĩcũgũ (two males and two females) and Ndia (one from each sex) only.

Graphemes <b> and <mb> for /h/ were presented by speakers from Ndia and Gĩcũgũ only. Grapheme <b> was produced by all the Gĩcũgũ respondents and <mb> was produced by one male and two females. Coming to Ndia, <mb> was produced by two male speakers and one female.

When the respondents read their translations, there were various surface realizations. For /β/, phones [β], [ɸ], [v] and [f] were realized. While [β] and [ɸ] were presented by respondents from all the dialects, [f] was given by respondents from all the dialects except Gĩcũgũ. The phone [v] was produced by two males and one female from Gĩcũgũ. One female from Ndia produced it and so did a male and two females from the Southern dialect. These responses are shown in Table 4.18 below.

For phoneme /ð/, respondents from all the dialects produced [ð]. Phone [^{n}d] was produced by two males and two females from Gĩcũgũ. It was generated in the word medial position.

Phoneme /ʃ/ was realized as [ʃ] by respondents from all the dialects. It was also realised as [s] by respondents from all the dialects except Mathĩra. Two males and one female from Mathĩra realised it as [tʃ]. The responses are also presented in Table 4. 18.

Phoneme /ɣ/ was realised as [ɣ] by respondents across the dialects. Two male respondents from Gĩcũgũ realised it as [k]. One male from Ndia as well as two females from Gĩcũgũ also realised it as [k]. This information is also shown in Table 4. 18.

Phoneme /h/ was presented as [b] by all the three males and two females from Gĩcũgũ. One female from Ndia also gave it as [b]. The phoneme was also vocalized as [^{m}b] by a male and a female from Gĩcũgũ, and a female from Ndia. We show these responses in Table 4. 18 below.

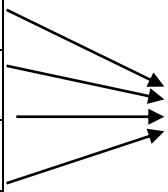
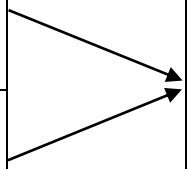
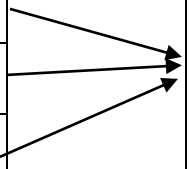
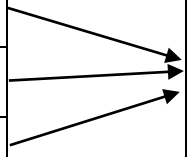
Table 4. 18: Respondents and Differing Fricative Phones

Phoneme/ Phone		Gĩcũgũ		Ndia		Mathĩra		Northern		Southern		**Total**	
		M	F	M	F	M	F	M	F	M	F	**M**	**F**
/β/	[f]	-	-	-	1	1	3	-	2	3	-	**4**	**6**
	[v]	2	1	-	1	-	-	-	-	1	2	**3**	**4**
/ð/	[^{n}d]	2	2	-	-	-	-	-	-	-	-	**2**	**2**
/ʃ/	[s]	1	2	2	2	-	-	-	1	2	1	**5**	**6**
	[tʃ]	-	-	-	-	2	1	-	-	-	-	**2**	**1**
/ɣ/	[k]	2	2	1	-	-	-	-	-	-	-	**3**	**2**
/h/	[β]	3	2	-	1	-	-	-	-	-	-	**3**	**3**
	[^{m}b]	1	2	-	1	-	-	-	-	-	-	**1**	**3**

Legend: **M** = males **F** = females

The information about the fricatives and their representation graphemically is summarized in Table 4.19 below.

Table 4. 19: The Gĩkũyũ Fricatives

Phoneme	Surface realization	Transition	Graphemic representation	Surface graphemic representations
/β/	[β]		<b>	<b>
	[ɸ]			<mb>
	[v]			<v>
	[f]			<f>
/ð/	[ð]		<th>	<th>
	[^{n}d]			<nd>
				<ny>
/ʃ/	[ʃ]		<c>	<c>
	[tʃ]			<ch>
	[s]			<s>
/ɣ/	[ɣ]		<g>	<g>
	[k]			<k>
/h/	[h]		<h>	<h>
	[β]			<b>
	[^{m}b]			<mb>

We observe from the table above that no phoneme has a single surface realisation. Additionally, each phoneme has several graphemic representations. Phoneme /β/ has three allophones and four allographs. /ð/ has two allophones and three allographs. /ʃ/ has three allophones and three allographs. /ɣ/ has two allophones and two allographs, while /h/ has three allophones and three allographs.

4.3.5 The Alveolar Tap /ɾ/

The respondents were asked to translate words with the phoneme and write them. Their graphemic responses are captured below.

(5)

Phoneme	Grapheme	Initial	Gloss	Medial	Gloss
/ɾ/	<r>	**r**ita	"litre"	me**r**ia	'swallow'
	<l>	**l**ita	"litre"	-	

In the table below, we show the number of respondents who produced the deviant grapheme <l>, their sex and dialect.

Table 4. 20: Respondents Producing Deviant <l> for <r>

Grapheme	Gĩcũgũ		Ndia		Mathĩra		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
<l>	1	2	1	1	2	1	1	2	-	-	**5**	**6**

Legend: **M** = males **F** = females

The grapheme was generated by respondents from all the dialects except the Southern. Also, there were more female respondents than male ones.

When the respondents read their translations, two phones, [ɾ] and the lateral [l], were produced. The two sounds are allophones in Gĩkũyũ and occur in free variation. Phone [l] was most clearly produced by three respondents from Mathĩra, one female from Northern, and a male from Southern. This is shown in the following table.

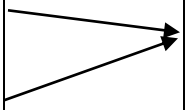
Table 4. 21: Respondents Producing Phone [l] for the Tap /ɾ/

Phoneme/ Phone	Gĩcũgũ		Ndia		Mathĩra		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
/ɾ/ [l]	-	-	-	-	1	2	-	1	1	-	**2**	**3**

Legend: **M** = males **F** = females

We note that though some Gĩcũgũ and Ndia respondents write <l> in place of <r>, they do not articulate the deviant phone [l]. It could be that they process /ɾ/ as <l> as a result of the allophonic nature of /ɾ/ and /l/ in Gĩkũyũ. On the other hand, a male from Southern who articulates /l/ writes it as <r>. The explanation possibly lies in his mastery of the graphemic system of Gĩkũyũ. The information given above is summarised in the following table.

Table 4. 22: The Alveolar Tap

Phoneme	Surface realization	Transition	Graphemic representation	Surface Graphemic representations
/ ɾ /	[ɾ]		<r>	<r>
	[l]			<l>

The table above shows that the phoneme has two allophones and two allographs.

4.3.6 The Approximants

Gĩkũyũ has two approximants, the bilabial approximant /w/ and the palatal approximant /j/**.** Each of the approximants has a single physical realization and a single grapheme representing it. We present data below to show how the graphemes were presented in initial and medial word positions.

(6)

	Phoneme	Grapheme	Initial	Gloss	Medial	Gloss
a)	/w/	<w>	**w**aku	'yours'	ta**w**a	'lamp'
b)	/j/	<y>	**y**ao	'theirs'	tha**y**ũ	'peace'

This data shows that no discrepancies were generated by the respondents.

When the respondents were asked to read the words with the phoneme, they produced a single phone for each phoneme as shown in the table below.

Table 4. 23: The Approximants

Phoneme	Surface realization	Transition	Graphemic representation	Surface Graphemic representations
/w/	[w]	⟶	<w>	<w>
/j/	[y]	⟶	<y>	<y>

From the table above, there were no discrepancies between the target graphemic representations and their surface representations.

4.3.7 The Gĩkũyũ Short Vowels and Graphemes

The data shown below, (7), was produced by the respondents for the vocalic graphemes in the word initial, medial and final positions.

(7)

	Phoneme	Grapheme	Initial	Gloss	Medial	Gloss	Final	Gloss
a)	/a/	<a>	**a**gũũta	‘lazy people’	y**a**o	‘theirs	ndereb**a**	‘driver’
b)	/ɛ/	<e>	**e**rĩ	‘two people’	nd**e**reba	‘driver’	onir**e**	‘s/he saw’
		<a>	-		nd**a**reba	‘driver’	-	
c)	/i/	<i>	**i**rio	‘food’	ir**i**o	‘food’	njong’**i**	‘ugly’
		<ĩ>	**ĩ**rio	‘food’	ir**ĩ**o	‘food’	njong’**ĩ**	‘ugly’

d)	/e/	<ĩ>	**ĩ**kumeria	‘will swallow’	k**ĩ**roto	‘dream’	ĩĩr**ĩ**	**‘**two’
		<i>	**i**kumeria	‘will swallow’	k**i**roto	‘dream’	ĩĩr**i**	‘two’
e)	/ɔ/	<o>	**o**nire	‘s/he saw’	ny**o**ka	‘snake’	ya**o**	**‘**theirs’
		<u>	-		ny**u**ka	‘snake’	-	
f)	/u/	<u>	**u**ma	‘move out’	nd**u**ka	‘shop’	nya**u**	**‘**cat’
		<ũ>	**ũ**ma	‘move out’	nd**ũ**ka	‘shop’	-	
g)	/o/	<ũ>	**ũ**mũthĩ	‘today’	bet**ũ**rũ	‘petrol’	thay**ũ**	**‘**peace’
		<u>	**u**mũthĩ	‘today’	bet**u**rũ	‘petrol’	thay**u**	‘peace’
		<o>	-		bet**o**rũ	‘petrol’	thay**o**	‘peace’

From the data, only the low vowel /a/ was represented with one grapheme and so it had no discrepancies. For the other phonemes, discrepancies were noted.

The mid [+ATR] /e...o/, in addition to being represented with their designed graphemes <ĩ> and <ũ> were represented with the designated graphemes for the high [+ATR] /i...u/ which are <i> and <u>, respectively. The converse also occurred where the high [+ATR] vowels were represented with their designated graphemes as well as those for the mid [+ATR] vowels.

In Table 4. 24 below, we show deviant graphemes against the phonemes they were generated for. We also show the number of respondents who presented those graphemes, their dialects and sex.

Table 4. 24: Deviant Graphemes, Number of Respondents by Dialect and Sex

Grapheme/ Phoneme	Gĩcũgũ		Ndia		Mathĩra		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F		F	**M**	**F**
/ɛ/ <a>	-	-	-	-	-	-	1	-	-	-	**1**	**0**
/i/ <ĩ>	-	-	1	1	-	1	-	-	-	1	**1**	**3**
/e/ <i>	3	3	2	1	2	2	3	2	3	2	**13**	**10**
/ɔ/ <u>	-	1	-	1	-	-	-	-	1	-	**1**	**2**
/u/ <ũ>	-	-	1	-	-	-	-	-	1	-	**2**	**0**
/o/ <u> <o>	3 3	3 3	3 2	3 2	1 -	3 -	3 -	2 1	3 1	1 -	**13** **6**	**12** **6**

Legend: **M** = males **F** = females

The front mid [-ATR] /ɛ/ is represented with the grapheme for the [Low] vowel /a/ by one male respondent from Northern dialect. The front high [+ATR] /i/ is represented with <ĩ>, the grapheme for the front mid [+ATR] /e/, by one male from Ndia, and a single female each from Ndia, Mathĩra and Southern. This is a

total of four (4) respondents. The front mid [+ATR] vowel /e/ is represented with <i>, the grapheme for the front high [+ATR], by a total of thirteen (13) males and ten (10) females from all the dialects. This is a prevalence of 86.6% for the males and 66.6% for the females. The back [-ATR] /ɔ/ is represented with the back high [+ATR] grapheme <u> by two females, each from Gĩcũgũ and Ndia, and one male from Southern. The back mid [+ATR] /o/ is represented with both <u> and <o>, the respective graphemes for the back high [+ATR] /u/ and the back mid [-ATR] /ɔ/. The representation with <u> is by thirteen (13) males or 86.6% and six (12) females or 80%. These respondents are spread over all the dialects. Grapheme <o> is generated by all the respondents from Gĩcũgũ, two thirds of both the males and females from Ndia, a third of the females from Northern, and a third of the males from Southern. The respondents from Mathĩra do not generate the grapheme.

In the reading task, the respondents produce various vocalic graphemes. These are shown in Figure 4.25 below. The figure also shows the graphemes representing the phones and phonemes.

Table 4. 25: Short Vocalic Phonemes and Graphemes

Phoneme	Surface realization	Transition	Grapheme	Graphemic surface realization
/i/	[i]	→	<i>	<i>
	[e]			< ĩ >
/e/	[e]	→	< ĩ >	< ĩ >
				< i>
/ɛ/	[ɛ]	→	<e>	<e>
				<a>
/a/	[a]	→	<a>	<a>
/ɔ/	[ɔ]	→	<o>	<o>
				<u>
/o/	[o]	→	<ũ>	< ũ >
				<u>
				<o>
/u/	[u]	→	<u>	<u>
				<ũ>

From the table above, each of the phonemes has two allographs, except for /o/ which has three. We observe that the front high and mid [+ATR] vowels interchange graphemes though the propensity is to interchange <i> for <ĩ>. This phenomenon might be explained by the fact that height may not be as decisive a factor in identifying the differences in the sets of the vowels. Rather, the 'voice quality' of being breathy or creaky may provide auditory differentiation. It may be that the discrimination in the graphemes is not held important by the speakers because, somehow, they differentiate the phonemes. For this reason, they interchange the graphemes. But this discrimination is important between mid [+ATR] and mid [-ATR] vowels. The two sets of vowels contrast (Peng, 2003). We note that the grapheme for the front mid [+ATR] vowel is not lowered to <e>. The back mid [+ATR] is lowered to <o> though it is also raised to <u>. We submit that respondents clearly discern the difference between [+ATR] and [-ATR] vowels and would wish to make the distinction clear, at least, orthographically. That also explains why the low front [-ATR] [ɛ] is lowered to [a] rather than raised to [e].

4.3.8 Gĩkũyũ Long Vowels

As noted in § 4.2.2 above, in Gĩkũyũ, vowel length is phonemic. In orthography, length in the initial position is not indicated. Respondents were asked to translate and write words which had long vowels in the initial position. The data below shows the graphemes they presented.

(8)

Phoneme	Grapheme	Initial	Gloss
a) /aː/	<a>	**a**nake	young men
	<aa>	**aa**nake	young men
b) /ɛː/	<e>	**e**rĩ	two
c) /iː/	<i>	**i**tũ	ours
	<ii>	**ii**tũ	ours
d) /eː/	<ĩ>	**ĩ**rĩ	two
	<ĩĩ>	**ĩĩ**rĩ	two
	<ii>	**ii**rĩ	two
	<i>	**i**ri	two
e) /ɔː/	<o>	**o**the	all
	<oo>	**oo**the	all
f) /uː/	<u>	**u**ngumania	corruption
	<uu>	**uu**ngumania	corruption
	<ũ>	**ũ**ngumania	corruption
	<i>	**i**ngumania	corruption
g) /oː/	<ũ>	**ũ**ĩ	knows
	<ũũ>	**ũũ**ĩ	knows
	<uu>	**uu**ĩ	knows

The data shows that in all instances, the designated grapheme is used. In all cases, vowel length is also indicated. For the front mid high [+ATR] /e:/, grapheme <i> which represents the front high [+ATR] /i/, was used and doubled as well. The back high [+ATR] /u:/ is represented with <uu> in addition to <i> and <ũ>. The grapheme <ũ> for the mid high [+ATR] is represented with <u> and <uu> in some cases. The grapheme <ii>, which indicates length, is also used. Additionally, /i/, the counterpart of the front high [+ATR] vowel /i:/, is produced.

The back mid high [+ATR] is represented with its grapheme <ũ> which is also doubled to <ũũ>. The grapheme for the high back [+ATR], <u> is used and also doubled in the representation.

In Table 4.26 below, we show the graphemes that differed from the ones in the graphemic inventory, either through doubling of graphemes or by use of an unrelated grapheme. We also show the number of respondents who generated these deviant vocalic graphemes, as well as their dialect and sex.

Table 4.26: Variant Initial Long Graphemes, Number of Respondents by Dialect and Sex

Phoneme/ Grapheme	Gĩcũgũ		Ndia		Mathĩra		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
/aː/ <aa>	-	-	-	-	-	-	1	-	-	-	**1**	**0**
/ɛː/ <ee>	-	-	-	-	-	-	1	-	-	-	**1**	**0**
/iː/ <ii>	2	-	-	1	-	-	1	-	-	2	**3**	**3**
/eː/ <ĩĩ>	-	-	-	1	-	-	1	-	-	1	**1**	**2**
<ii>	-	-	-	-	-	-	1	-	-	-	**1**	**0**
/ɔ/ <oo>	-	-	-	1	-	-	1	-	-	1	**2**	**1**
/uː/ <uu>	--	-	-	-	-	-	1	-	-	1	**1**	**1**
<ũ>	-	-	-	1	-	-	-	-	-	-	**0**	**1**
<i>	-	1	-	-	-	-	-	-	-	-	**-**	**1**
/o/ <ũũ>	-	-	-	1	-	-	-	-	-	-	**0**	**1**
<uu>	-	-	-	-	-	1	-	-	-	-	**0**	**1**
<u>	-	1	-	-	-	-	-	-	-	-	**0**	**1**

Legend: **M** = males **F** = females

From the table, we notice that although all the phonemes had variant representations, /aː/, /ɛː/, /iː/ and /ɔː/ had one each, /eː/ had two, while /uː/ and /oː/ had three each. Respondents from Mathira used the correct grapheme and so did all the males from Ndia and Southern, as well as all the males from Northern.

In the reading task, the graphemes that were produced are shown in Table 4.27.

Table 4.27: The Gĩkũyũ Long Vowels in the Initial Word Position

Phoneme	Surface realization	Transition	Grapheme	Graphemic surface realization
/aː/	[aː]	→	<a>	<a> <aa>
/ɛː/	[ɛː]	→	<e>	<e> <ee>
/iː/	[iː]	→	<i>	<i> <ii>
/eː/	[eː]	→	< ĩ>	< ĩ> < ĩĩ> <ii> <i>
/ɔː/	[ɔː]	→	<o>	<o> <oo>
/uː/	[uː]	→	<u>	<u> <uu> < ũ> <ı> < ũu>
/oː/	[oː]	→	<ũ>	< ũ> <ũũ> <uu>

From the table, we note that graphemes in the initial position are doubled. This is inconsistent with Gĩkũyũ orthography. We also note that where other graphemes are used, they are invariably used for vowels sharing the same [ATR] feature. In one case of discrepancy, there is fronting of /u:/ to /i/ where the two are [+ATR]. This strengthens our argument that [+ATR] graphemic representations do not interchange with [-ATR] representations.

In the medial and final positions, vowel length is graphemically represented with doubled graphemes. The respondents gave the following responses for the long vowels in these positions.

(9)

	Phoneme	Grapheme	Medial	Gloss	Final	Gloss
a)	/a:/	<aa>	mat**aa**ro	‘advice’	mũg**aa**	**‘**acacia’
		<a**>**	mat**a**ro	‘advice’**-**		
b)	/ɛ:/	<ee>	mab**ee**be	‘maize stalks’	at**ee**	‘s/he lost’
		<e>	mab**e**be	‘maize stalks’	at**e**	‘s/he lost’
		<ea**>**	**-**		at**ea**	**‘**she lost’
c)	/i:/	<ii>	r**ii**ko	‘kitchen	mũnab**ii**	‘prophet’
		<i>	r**i**ko	‘kitchen’	**-**	

d)	/e:/	<ĩĩ>	n**ĩĩ**kameria	‘it will swallow	kah**ĩĩ**	‘small boy’
		<ĩ>	n**ĩ**kameria	‘it will swallow’	-	
		<ii>	n**ii**kameria	‘it will swallow’	kah**ii**	‘small boy’
		<i>	n**i**kameria	‘it will swallow’	-	
		<ĩi>	-		kah**iĩ**	‘small boy’
e)	/ɔ:/	<oo>	kĩr**oo**to	‘dream’	ndir**oo**	**‘**drawer’
		<o>	kir**o**to	‘dream’	ndir**o**	‘drawer’
f)	/u:/	<uu>	h**uu**hu	‘bats’	kab**uu**	‘grey’
		<u>	h**u**hu	‘bats’	-	
		<ũ>	h**ũ**hu	‘bats’	-	
		<ũu>	-		kab**ũu**	‘grey’
g)	/o:/	<ũũ>	ag**ũũ**ta	‘lazy people’	-	
		<ũ>	ag**ũ**ta	‘lazy people’	-	
		<u>	ag**u**ta	‘lazy people’	-	

As was the case with length in the initial position, /a:/, /i:/ and /ɔ:/ record their respective single graphemes and then the graphemes are doubled.

Phoneme /ɛ:/ records its double grapheme <ee> and the single representation too <e>. It also records a representation with <ea>. This shows a case of a glide from the slightly high position of low [-ATR] to the low/ neutral position of /a/. This is the same case observed in the response given in (9 b) above.

In the reading task, the phones that were recorded are shown in Table 4.28. The table also shows the graphemes that were presented.

Table 4.28: The Gĩkũyũ Long Vowels in Medial and Final Positions

Phoneme	Surface realization	Transition	Grapheme	Graphemic surface realization
/aː/	[aː]	→	<aa>	<aa> <a>
/ɛː/	[ɛː]	→	<ee>	<ee> <e> <ea>
/iː/	[iː] [i]	→	<ii>	<ii> <i>
/eː/	[eː] [e]	→	<ĩĩ>	<ĩĩ> <ĩ> <ii> <i>
/ɔː/	[ɔː]	→	<oo>	<o> <oo>
/uː/	[uː]	→	<uu>	<uu> <u> <ũũ> <ũ> <i>
/oː/	[oː]	→	<ũũ>	<ũũ> <ũ> <uu> <u>

Benson (1964) observes that vowels in non-initial word position are invariably long before the nasalized plosives and that initial vowels are invariably short unless they have undergone coalescence. From the data presented above, phonemes /a:/, /i:/ and /ɔ:/ are represented with either their doubled or single grapheme. This in itself causes discrepancies. The mid [+ATR] vowels /e: o:/ have their graphemes doubled (<ĩĩ>, <ũũ>) or single (<ĩ>, <ũ>). In addition, they take the high [+ATR] symbol and also double it (<i> and <ii> and <u> and <uu>). We observe here that the representation for [+ATR] vowels does not mix with that of [-ATR] vowels.

The grapheme for the front [-ATR] /ɛ:/ is double <ee> and in other cases single <e>. It also appears as a combination of the single grapheme and the grapheme for the low vowel <a>. Here, we note preference for tongue retraction rather than advancement. As observed in the paragraph above, the [+ATR] graphemic representation does not mix with the [-ATR] representation.

4.3.9 Vowels in Concatenation

There are twenty-four possible vocalic graphemic combinations in Gĩkũyũ (Armstrong, 1967, p. 2). These are /ie/, /iɛ/, /ia/, /iɔ/, /io/, /iu/, /ei/, /ɛi/, /ai/, /oi/, /ɔi/, /eɛ/, /ea/, /eɔ/, /eo/, /oe/, /ɔe/, /ue/, /oɛ/, /uɛ/, /oa/, /ua/, /uɔ/, and /uo/. Their graphemes are <iĩ>, < ie>, <ia>, <io>, <iũ>, <iu>, <ĩi>, <ei>, <ai>, <ũi>, <oi>, <ĩe>, <ĩa>, <ĩo>, <ĩũ>, <ũĩ>, <oĩ>, <uĩ>, <ũe>, <ue>, <ũa>, <ua>, <uo>, and <uũ>, respectively. However, a close examination of the language reveals that

there are ten more possible combinations which neither Armstrong nor any other linguist, to the best of our knowledge, has hitherto identified. These are /eu/, /ɛɔ/, /ɛa/, /ou/, /ɔɛ/, /ɔa/, /aɔ/, /ao/, /au/ and /ae/. We present these phonemes and their graphemes, together with examples of words, in the table below.

Table 4.29: Vowels in Concatenation

Phoneme	grapheme	word	gloss
/eu/	<ĩu>	mũr**ĩu**	drunkard
/ɛɔ/	<eo>	kĩh**eo**	gift
/ɛa/	<ea>	theb**ea**	belch
/ou/	<ou>	k**ũu**	there
/ɔɛ/	<oe>	**mũ**tong'oe	tail
/ɔa/	<oa>	ng'**oa**	assemble
/aɔ/	<ao>	**ao**	theirs
/ao/	<aũ>	t**aũ**ra	translate
/au/	<au>	h**au**	there
/ae/	<aĩ>	m**aĩ**	water

These phonemes and graphemes were analysed alongside the others. We, however, noted that it is not possible to identify Gĩkũyũ words with these concatenated phonemes in the word initial position.

To generate data, respondents were served with English carrier words that they were to translate into Gĩkũyũ. The phonemes were in the word initial, medial and final positions, where possible. We present their responses below. We group the responses according to where the movement originates from.

(10) From Front High [+ATR] /i/

	Phoneme	grapheme	initial	gloss	medial	gloss	final	gloss
a)	/iu/	<iu>	-		c**iu**go	'words'	-	
b)	/ie/	<iĩ>	-		c**iĩ**ko	'actions'	mũc**iĩ**	'home'
		<ii>	-		c**ii**ko	'actions'	mũc**ii**	'home'
c)	/io/	<iũ>	-		c**iũ**ria	'questions'	rũc**iũ**	'tomorrow'
		<iu>	-		c**iu**ria	'questions'	-	
		<ĩũ>	-		c**ĩũ**ria	'questions'	-	
		<io>	-		c**io**ria	'questions'	-	
		<ĩu>	-		c**ĩu**ria	'questions'	-	
d)	/iɛ/	<ie>	-		c**ie**ro	'thighs'	njar**ie**	'I search'
e)	/iɔ/	<io>	-		c**io**the	'all'	ir**io**	'food'
f)	/ia/	<ia>	-		c**ia**na	'children'	uungumar**ia**	'corruption'
		<ĩa>	-		c**ĩa**na	'children'		

In the data presented above, it was not possible to get carrier words with the phonemes in the word initial position. It was also not possible to get a word with /iu/ in the word final position. In all the cases, the representation for /i/ is with the correct grapheme, <i> except in the instance where <ĩ> is used. This is a situation where there is a process of vowel lowering from high to mid but with maintenance of the [+ATR] characteristic.

A summary of the discrepancies is given in Table 4.30 below.

Table 4. 30: Front High [+ATR] /i/ Variant Graphemes against Number of Respondents

Phoneme/ Grapheme	Gĩcũgũ		Ndia		Mathĩra		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
/ie/<ii>	2	-	-	-	-	-	-	1	-	-	**2**	**1**
/io/<iu>	2	0	3	0	1	1	2	2	2	1	**10**	**4**
<ĩũ>	-	-	-	-	1	-	1	-	-	-	**2**	**0**
<io>	1	-	-	-	-	-	-	-	-	-	**1**	**0**
<ĩu>	-	-	-	-	-	-	-	-	1	-	**1**	**0**
/ia/<ĩa>	1	-	-	-	-	-	-	1	1	-	**2**	**1**

Legend: **M** = males **F** = females

From the table, we notice that the movement from the Front High [+ATR] /i/ to Back Mid [+ATR] /o/ generated the most discrepancies. The representation of /o/ with /u/ was by respondents across all the dialects and by a majority of the males (66.6 %).

(11) From Front Mid [+ATR] /e/

	Phoneme	grapheme	initial gloss	medial	gloss	final	gloss
a)	/ei/	<ĩi>	-	th**ĩi**niĩ	'inside'	gĩth**ĩi**	'mill'
		<ii>		th**ii**niĩ	'inside'		
b)	/eu/	<ĩu>	-	-		mũr**ĩu**	'drunkard'
c)	/eo/	<ĩũ>	-	k**ĩũ**ria	'question'	-	
		<ĩu>		k**ĩu**ria	'question'		
		<iu>		k**iu**ria	'question'		
d)	/eɛ/	<ĩe>	-	m**ĩe**ri	'months'	ar**ĩe**	'to let eat'
		<ie>		m**ie**ri	'months'	ar**ie**	'to let eat'
e)	/eɔ	<ĩo>	-	k**ĩo**ro	'latrine'	nĩk**ĩo**	'that's why'
		<io>		k**io**ro	'latrine'		
f)	/ea/	<ĩa>	-	m**ĩa**ka	'years'	r**ĩa**	'eat'

<ia> m**ia**ka ‘years’ r**ia** ‘eat’

From the data above, we observe that in the word final position, there is no discrepancy except with /ea/. Most discrepancies occur in the word medial position. A summary of this information is given in Table 4.31 below.

Table 4. 31: Front Mid [+ATR] /e/, Variant Graphemes against Number of Respondents

Phoneme/grapheme	Gĩcũgũ		Ndia		Mathĩra		Northern		Southern		**Total**	
	M	F	M	F	M	F	M	F	M	F	**M**	**F**
/ei/ <ii>	1	-	1	1	-	-	-	1	-	-	**2**	**2**
/eu/<iu>	-	-	-	-	-	-	-	1	1	-	**1**	**1**
/eo/<iu>	1	-	-	-	-	1	2	1	-	1	**3**	**3**
<iũ>	-	1	-	-	-	-	-	-	1	-	**1**	**1**
<ĩu>	-	-	-	-	-	1	-	-	-	-	**-**	**1**
/eɛ/ <io>	2	-	2	1	2	1	1	1	1	-	**8**	**3**
/ea/ <ia>	2	2	1	1	2	-	2	1	-	1	**7**	**5**

Legend: **M** = males **F** = females

(12) From Front [-ATR] /ɛ/

	Phoneme	Grapheme	Initial	Medial	Gloss	Final	Gloss
a)	/ɛi/	<ei>	-	t**ei**thia	‘help’	kang’ei	‘middle aged woman’
b)	/ɛɔ/	<eo>	-	-		kĩh**eo**	‘gift’
c)	/ɛo/	<eũ>	-	-		mĩ**eũ**	‘yawn’(n)
d)	/ɛa/	<ea>	-	ah**ea**ga	‘gives’	theb**ea**	‘belch’

There were no discrepancies in the generation of the graphemes.

(13) From [+Low] /a/

	Phoneme	Grapheme	Initial	Gloss	Medial	Gloss	Final	Gloss
a)	/ai/	<ai>	**ai**ru	‘co-wives’	m**ai**tho	‘eyes’	ig**ai**	‘share’(n)
b)	/ae/	<aĩ>	-		g**aĩ**ra	‘share out’	m**aĩ**	‘water’
c)	/aɔ/	<ao>	-		-		wao	‘theirs’
d)	/ao/	<aũ>	**aũ**ra	‘relieve’	t**aũ**ra	‘interpret’	gĩt**aũ**	‘speak plainly’
e)	/au/	<au>	-		t**au**rũ	‘towel’	ny**au**	‘cat’

The generation of these graphemes also occasioned no discrepancies.

(14) From Back [-ATR] /ɔ/

Phoneme	Grapheme	Initial	Gloss	Medial	Gloss	Final	Gloss

a) /ɔi/ <oi> **oi**ga ‘says’ tor**oi**ra ‘loiter’ ng**oi**‘baby-carrier’

b) /ɔe/ <oĩ> **oĩ** ‘knows’ nd**oĩ**kaine ‘he’s unknown’nd**oĩ**‘he doesn’t know’

c) /ɔɛ/ <oe> - h**oe**ra ‘pray for’ mũtong’**oe** ‘tail’

d) /ɔa/ <oa> - - - ng’**oa** ‘assemble’

In the presentation of these graphemes, we also observe that there were no discrepancies.

(15) From Back Mid [+ATR] /o/

Phoneme	Grapheme	Initial	Gloss	Medial	Gloss	Final	Gloss
a) /ou/	<ũu>	-		-		k**ũu**	‘there’
b) /oi/	<ũi>	**ũi**ru	‘jealousy’	m**ũi**ko	‘cooking stick’	kany**ũi**	‘razor blade’
	<oi>	-		m**oi**ko	‘cooking stick’	kany**oi**	‘razor blade’
	<i>			m**i**ko	‘cooking stick’		
c) /oe/	<ũĩ>	**ũĩ**kaine	‘known’	g**ũĩta**	‘call’	nj**ũĩ**	‘rivers’
	<ui>	**ui**kaine		-		nj**ui**	‘rivers’
	<uĩ>	**uĩ**kaine		-		-	
d) /oɛ/	<ũe>	-		r**ũe**nji	‘traditional razor’	nj**ũe**	‘peeled’(adj)
e) /oa/	<ũa>	-		g**ũa**ka	‘to build’	ih**ũa**	‘flower’
	<ua>	-		g**ua**ka	‘to build’	ih**ua**	‘flower’

There is no movement from this phoneme to the back mid [-ATR] vowel and the movement to the front mid [-ATR] occasions no discrepancies. The movement to back high [+ATR] /u/ also has no discrepancies. This can be explained by the principle of least effort.

(16) From Back High [+ATR] /u/

	Phoneme/Grapheme		Initial	Medial	Gloss	Final	Gloss
a)	/ui/	<ui>	-	iny**ui**ki	‘you (pl) alone’	ng**ui**	‘dog’
b)	/ue/	<uĩ>	-	nj**uĩ**rĩ	‘hair’	ith**uĩ**	‘we’
c)	/uɛ/	<ue>	-	iny**ue**rĩ	‘both of you’	ng**ue**	‘I die’
d)	/ua/	<ua>	-	h**ua**na	‘look like’	k**ua**	‘die’
e)	/uɔ/	<uo>	-	g**uo**ko	‘arm’	ng**uo**	‘dress’
f)	/uo/	<uũ>	-	-		gĩk**uũ**	‘death’

There were no carrier words with the phonemes in the initial position. In the presentation of the graphemes, no discrepancies were noted.

The information presented in §4.3.1 to § 4.3.9 is on the Gĩkũyũ consonantal and vocalic phonemes. Also included are the graphemes in the Gĩkũyũ graphemic inventory. The graphemes generated by the respondents are also given as well as the phones produced for each phoneme. We have established that there exist discrepancies between the graphemic inventory and some of the graphemes that were generated. In the following section, we look at the principles that underlie an

orthography and examine how they are observed or contravened by both the Gĩkũyũ orthography designers and the Gĩkũyũ speakers. We also identify Optimality Theory Constraints from the principles and use them to analyse the Gĩkũyũ graphemes generated in the study.

4.4 The Orthography Criteria and the OT Constraints

There are many criteria that are used in setting up an orthography. They include phonological adequacy, simplicity, areal appropriateness, community ownership, accuracy, consistency, convenience and harmonization (Bradley, 2003; Malone, 2004; Stark, 2010). Viewed together, these criteria deal with the representation of the sound system of the language, the ease with which the representation is used, and the attitude of the community towards that representation. With this in mind, we have condensed these criteria into three. These are the phonological criterion which encompasses phonological adequacy, accuracy and consistency; the simplicity criterion which subsumes areal appropriateness and convenience; and the community ownership criterion which incorporates harmonization. Each criterion gives rise to a number of principles. Baroni (2016) observes that these principles are based on perceptual, articulatory and psychological factors. The principles do conflict where the observance of one principle may of necessity conflict with the observance of one or more other principles. The conflicts arise from the fact that the principles are built on universal constraints which are grounded either on phonetics (because of perceptual and articulatory factors) or on cognition (because of psychological factors).

There are two types of constraints: Markedness and Faithfulness constraints. The constraints are constraint families rather than just constraints. These constraints are ranked differently by different languages.

Markedness constraint are based on the premises that all types of linguistic structures consist of 'marked' and 'unmarked' values. The unmarked values are cross-linguistically preferred and basic in all grammars, while the marked values are cross-linguistically avoided and used by grammars only to create contrast (Kager, 1999, p. 2). Markedness constraints require that the output will have some 'optimal' shape (van Oostendorp, 2004) that corresponds to the input. In orthography, they 'avoid excessive effort in production and lack of salience in perception' (Baroni, 2016, p. 293). They require that contrasts be maintained, and that they be perceptually distinct. Flemming (2002) and Padgett (1997, 2003) identify two basic, and conflicting, markedness imperatives that drive phonology. These are to *maximize* the perceptual distinctiveness between contrasting forms and to *minimize* effort (e.g., articulatory, processing). Simply put, markedness constraints maximize perceptual distinctiveness and minimize articulatory or processing effort and ensure that the phonotatics of the language are maintained.

Faithfulness constraints require that the output does not differ markedly from the input. They protect the input from undergoing changes and ensure that the input is, in a way, the same as the output. This is because although the input and the

output are not qualitatively the same entity, they are the same types of objects in reduplication.

In the following analyses, we present the orthography criteria and the principles under each criterion. We label the principles P*n* where 'P' stands for the principle and '*n*' for the number of the principle. We also present the Universal Constraints that each principle is built upon.

4.4.1 The Phonological Adequacy Criterion

Phonological adequacy takes into cognisance the idea that an orthography should fit the language system. It is thus concerned with the graphemic representation of phonological features. Taken into consideration is how phonological processes might influence the representation of the phonemes. Such processes include gliding, assimilation and dissimilation, nasalization and vowel length. Suprasegmental features such as tone and stress may also be candidates for representation. In this study, the linguistic factors under consideration include nasalization, gliding and vowel length.

An important factor in the phonological adequacy is the functional load of a phoneme. Seifart (2006) expresses functional load as the importance a feature may have towards reading and writing in a given language. Karan (2006) further expresses it as the potential for confusion that may arise if the phoneme is not represented in a unique way. The functional load has a bearing on what phonological features will be represented by a given grapheme. Vowel length and

nasalization are two phonological features whose functional load impacts on graphemic representation in Gĩkũyũ. For example, not representing the nasal segment may cause confusion between the words '*mbathi*' (bus) and '*bathi*' (iron box); while, non-representation of vowel length can interfere with meaningful reading or writing of the minimal pair '*hota*' (be able) and '*hoota*' (win).

The frequency with which a phoneme occurs in a language will determine whether it has a heavy load or not. Some phonemes in Gĩkũyũ have a high frequency and so a heavy functional load as compared to others. For example, /j/ <y> in the initial word position occurs in only 17 words (Benson, 1964, pp. 561 – 562), as compared to /h/, <h> which occurs in over 772 words in the same position (pp. 134-180).

There are several principles that emanate from the phonological adequacy criterion.

P1. The first principle is the One phoneme - One grapheme - Principle. This principle requires that a single phoneme is represented by a single grapheme. A grapheme may be a single symbol or more than one symbol as is seen in<b> and <mb> in Gĩkũyũ. This principle heightens accuracy whereby, if there are X number of phonemes, there would be an identical number of graphemes. It also militates against ambiguities that may result from allophonic representations. An example where accuracy can be interfered with and ambiguity be created is in the representation of the palatal fricative /ʃ/ represented as <c> in Gĩkũyũ. Since /ʃ/

has the allophones [s] and [tʃ], if <s> and <ch> were also used for the phoneme, different interpretations would be given to a single phoneme. For instance, it would be unclear what the word ***c**ai* 'tea' means if we also had ***s**ai* and ***ch**ai.*

From P1, we can posit the following faithfulness constraints:

MAX- (P-INPUT, O-OUTPUT).

These constraints demand that segments in the input must be preserved in the output. They mitigate against deletion of a segment from the input. Thus, if the input was /ɔ:/, and grapheme <o> was proffered by respondents, the grapheme would be bereft of the segment [LENGTH] which would violate the constraint. However, this is a low ranked faithfulness constraint in Gĩkũyũ. In this study, the constraint is presented as:

MAX - (1PHONEME,1GRAPHEME) = MAX-(1P, 1G)

Pre-nasality is another segment that must not be deleted. A universal markedness constraint *NC-(NASAL, CONTINUANT) requires that nasals must not be followed by continuants. In Gĩkũyũ, all voiced plosives and affricates must be preceded by a nasal. In this study, this constraint can be rephrased thus:

NC- (NASAL, PLOSIVE) = NC - (N, P)

P2. A second principle emanating from the phonological adequacy criterion is the Consistency Principle. This principle requires that each orthographic symbol

represents one sound. This state is otherwise referred to as the grapheme-phoneme correspondence. The relationship between a grapheme and a phoneme should be regular, otherwise inconsistency and unpredictability will be created (Sasaki, 2003). Karan (2006) notes that the use of one grapheme to represent two or more phonemes constitutes underrepresentation. Graphemes that represent more than one symbol create ambiguity and impinge on accuracy. For example, <g> should only represent the velar fricative /ɣ/ and not any other phoneme such as the prenasalised velar plosive /ᵑɡ/.

Consistency also means that once a system or a symbol has been assigned, then it should be invariably used henceforth. There should be no replacement of the system or the grapheme with another. To give an example, in Gĩkũyũ orthography, a long vowel in the word initial position should invariably be represented by a single grapheme. Thus, /a:nakɛ/ (young men) should be written as <anake> and not <aanake>. If two graphemes are used, the principle will have been contravened. In this case, a conflict with P1 is noted.

A faithfulness constraint that is noted here is MAX-(P-OUTPUT, I-INPUT). In orthography, this can be restated as:

MAX-(1GRAPHEME, 1 PHONEME) = MAX - (1G, 1P)

Another relevant faithfulness constraint is IDENT- (FEATURE) that advocates one grapheme one phoneme consistency. We present it in orthography as:

IDENT-(GRAPHEME, PHONEME) =IDENT-(1G, 1P)

Another faithfulness constraint that is relevant from this principle is DEP-(P-INPUT/P-OUTPUT, O-OUTPUT). The constraint requires that all segments in the output must be present in the input. It bans the insertion or addition of a segment to the input which would thereafter be reflected in the output.

In this study, we present the constraint as:

DEP - (PHONEME, GRAPHEME) = DEP - (P, G)

In the constraint above, P is the phoneme and G the grapheme that carries all the segments in the phoneme.

P3. The third principle is the Write as You Speak Principle. This principle requires that each phone must be represented graphemically. The principle applies in a shallow orthography and is in direct contrast to P1. It takes care of phones rather than phonemes. Guerin (2008) observes that such a representation eases reading and spelling especially for the beginning reader and the non-fluent speaker, and that it is ideal to a language with no dialectal variation. In such a situation, where allomorphs occur, they may be represented separately. As an example, since in Gĩkũyũ some speakers physically articulate both [r] and [l] for the alveolar tap /ɾ/, both phones would be represented graphemically as <r> and <l>, respectively.

Several constraints can be posited here. A faithfulness constraint is IDENT-(FEATURES/PHONE, GRAPHEME).

In orthography terms, this can be presented as:

IDENT- (FEATURES,PHONE) = IDENT- (F, P)

This is a very low - ranking constraint owing to the fact that writing does not always reflect speaking. The representation of allophones bears this out and so does the representation of prosodies in most languages.

The markedness constraints that can be invoked here are VISIBILITY. This would refer to the requirement that, for example, salient features are represented, or vocalic length should be represented. In orthography terms, the constraint for the visibility of features can be given as:

VIS-(F)

The constraint for the visibility of vowel length can be given as:

VISIBILITY- (V-LENGTH) = VIS - (V-L)

Another markedness constraint would be DIACRITICS. This is the requirement that salient features beyond the segment, such as nasality, be represented. As an example, the diacritic (’) should differentiate /ᵑg/ from /ŋ/. The orthography presentation of the constraint can be:

DIACRITICS - (FEATURES) = DIA-(FT)

P4. The Convey Meaning Principle is another principle under the adequacy criterion. This principle is based on the fact that the primary aim of an orthography is to convey meaning (Stark, 2010). This hinges on the functionality of an orthography. One way of conveying meaning is by enhancing differentiation and, at the same time, minimizing distortion (Boersma, 1998). The orthography should be read or written in such a way that the least effort is used to convey the most information in the clearest way possible so that the intended meaning is not missed; the wrong meaning is not construed; and neither is there any struggle to decipher the intended meaning.

We propose a markedness constraint here:

CM - (CONVEY MEANING) = CM - (MEAN)

The constraint bans the writing of graphemes that do not convey meaning, or convey a different meaning in Gĩkũyũ. This is a very high ranking constraint.

P5. Another principle under this criterion is the Do Not Change the Inventory of Letters Principle (Wiese, 2004). This principle argues against the introduction of graphemes that do not belong to the graphemic inventory of a given language. Thus, in Gĩkũyũ, we may not have representations such as <d>, <j>, <f>, <v>, <p>, <s>, <z>, <q> and <x>.

From this principle, we posit a very high ranking markedness constraint which we present as:

INVENTORY-(GRAPHEME) = INV-(G)

4.4.2 The Simplicity Criterion and the Graphemes

This criterion gives rise to two principles which are explained below.

P6. This is the Principle of Overall Least Effort (Guerin, 2008, p. 59). The principle is based on the function that an orthography is meant to serve. This is whether it is for reading or writing, for beginning writers/readers or advanced ones, or whether it is for first language speakers or for non-fluent speakers. The principle requires that familiar symbols should not be applied in unfamiliar ways. Familiar symbols are those that are used by neighbouring languages that speakers are aware of. Speakers may hold a neighbouring language positively and would want to use the same graphemes. This principle applies in our study where English and Kiswahili are held in high esteem (Nabea, 2009) and are the immediate neighbours of Gĩkũyũ.

The faithfulness constraint invoked here is the IDENT-(FAMILIAR, UNFAMILIAR) in which familiar symbols should not be used unfamiliarly. We restate the constraint here in the negative as:

*IDENT-(FAMILIAR, UNFAMILIAR) = *IDENT- (F, U)

Two markedness constraints can be identified. The first advocates for simplicity and will be given in the negative as *COMPLEX (Wiese, 2004). In the analysis, we refer to it as:

*COMP

The second constraint also advocates writing in continuous flow given the fact that stopping to include diacritics such as the macron and apostrophes, requires more effort. The relevant constraint in our study is given in the negative as

*DIACRITICS = *DIA

P7. This is the 'Easiest for Users' principle. It combines salience in visual appearance and ease in writing. The principle requires that graphemes should be easy to read as well as easy to write by hand, or easy to type. Single symbols are easier to read than digraphs which are in turn easier to read than symbols with diacritics. For instance, it would be easier to read or write <g> than <ng> which in turn would be easier than <ng'>. A diacritic employed in an orthography, the apostrophe < ' > or the tilde < ˜ > for instance, are often abandoned in spontaneous writing (Coulmas, 1989 in Baroni, 2016). Symbols that are too much alike as well as symbols that are a crowding of adjoining letters are not easy to use. As an example, <u> and <ũ> in concatenation may prove difficult to use due to their visual similarity. It is noted that though unusual graphemes can be written using computer today, it may still be cumbersome and common keyboards may not have them. Such would include <ĩ> and <ũ>. Additionally, the use of computers might not be that widespread and printers may not be readily available. The writing of complicated graphemes by hand is also a daunting task. The

constraints evoked here are the same as those for P6, the *COMPLEX and *DIACRITICS

4.4.3 Community Ownership Criterion and the Constraints

This criterion arises from the realization that an orthography is designed with a community in mind and that it is futile to design an orthography which speakers are unwilling to use. Based on this, two principles are identified.

P8. One of the principles is the Familiarity Principle (Stark 2010; Williamson, 1984, pp. 10-11; Barnwell, 1998, p. 74). The orthography should match the speakers' expectations as to what their orthography should look like. The principle is based on the resultant feeling of correctness that speakers have when they are accustomed to their orthography. Additionally, they are aware of the orthographies of the neighbouring and related languages and the correlations and differences with theirs. This principle augurs well for standardization of the orthography but not for its reform.

The constraints evoked here are for faithfulness and can be given as FAMILIARITY. The constraints rank very low in Gĩkũyũ phonotactics. In this analysis, they are presented as:

FAM

P9. The second principle addresses issues to do with acceptability. The writing system should be acceptable to the speakers. They should not reject it on grounds

of its being non-representative or being too difficult, or even being impractical. Speaking on Gĩkũyũ, Githiora (2003) notes that there should be popular will by the users of a language to accept the orthography set for them.

From the principle, we can derive the low - ranking faithfulness constraints of ACCEPTANCE. These are presented in the analyses as:

ACCEPT

4.5 The Principles, Constraints and the Graphemes

Both markedness and faithfulness constraints work together to provide the orthography of Gĩkũyũ. Whereas markedness pays attention to the output form, faithfulness constraints focus on both input and output (Kager, 1999). The language ranks these constraints in a strict dominance hierarchy. The lower ranked constraints may be violated by the observance of higher ranked ones.

Typically, one or more markedness constraints are ranked above one or more faithfulness constraints. This means that when a conflict occurs, the markedness constraints will reign supreme.

In the sections below, we analyse each of the phonemes (the input) and its variant graphemes (all of which constitute the candidates) against the constraints. The input are cognitive entities while the output are physical entities - graphemes. In OT, competing candidates are pitted against constraints in tableaux. The constraints are given in the topmost row with the highest ranked constraint

occupying the most left slot. The candidates are presented in the left-most column. A constraint violation will be shown by an asterisk (*) against the candidate. A fatal violation that eliminates a candidate will be shown by an asterisk plus an exclamation mark (*!). The winner is the candidate who does not violate a higher ranking constraint as opposed to the other candidate(s) who violates a higher ranking constraint. This winner is shown with a pointing finger, thus ☞ (see § 3.2).

4.5.1 The Plosives and the Affricate

In Gĩkũyũ, the voiceless plosives, /t/ and /k/ are realized on the surface as [t] and [k] (see Figure 4.3). There are no discrepancies in their graphemic representations which are <t> and <k> respectively. This is characteristic of a shallow orthography (see § 5.1 below).

However, for the voiced bilabial plosive /^{m}b/ that should be represented by <mb>, and the voiced alveolar plosive /^{n}d/ that should be represented by <nd>,another additional grapheme is presented, <b> and <d>, respectively. The voiced velar plosive /^{ŋ}g/ whose grapheme should be <ng> has some one more grapheme representing it, that is <g>. The voiced plosives differ from their voiceless counterparts, not just in voice, but mainly in that they are preceded by a placeless nasal segment (see § 4.3.1). The voiced affricate /ɟ/ is also preceded by the placeless nasal. The phoneme is realised on the surface as [^{n}ɟ] and is represented in our data with two graphemes (see § 4.3.3).

Additionally, each of the pre-nasalised plosives had two surface realizations. These were [^{m}b] and [b] for the bilabial plosive /^{m}b/, [^{n}d] and [d] for the alveolar plosive /^{n}d/, [^{ŋ}g] and [g] for the voiced velar plosive /^{ŋ}g/ and [^{ɲ}ɟ] and [ɟ] for the palatal affricate /^{ɲ}ɟ/.

The use of <b> for /^{m}b/, <d> for /nd/, <g> for /^{ŋ}g/ and <j> for /^{ɲ}ɟ/ contravene P1, the One phoneme - One grapheme Principle. Though these graphemes are representative of the plosive segment of the phoneme, they leave the nasal segment unrepresented. In this case, we posit the faithfulness constraints MAX-(1P, 1G).

Since /^{m}b/ is prenasalised, we propose the markedness constraint:

NC- (NASAL, PLOSIVE) = NC- (N, P)

The representation of /^{m}b/ with <b>, and of /^{ŋ}g/ with <g>is a contravention of P2, the Consistency Principle. P2 is the grapheme-phoneme correspondence which requires that each graphemic symbol should represent one phoneme and that a grapheme should represent only one phoneme. For this reason, we invoke the faithfulness constraint IDENT- (1G, 1P). In the Gĩkũyũ orthography, grapheme <b> represents /β/ while <g> represents /ɣ/. Consequently, their use here means they each represent two phonemes now. This creates underdifferentiation between phonemes /^{m}b/ and /β/ on one hand, as well as between /^{ŋ}g/ and /ɣ/ on the other. It also makes the orthography deep (see § 5.1 below).

We also derive the DEP- (P-INPUT/P-OUTPUT, O-OUTPUT) constraints from this principle. The representation of /^{n}d/ with <d> means that a new segment has been added. We present the constraint here as DEP- (P, G)

P4, the Convey Meaning Principle is also contravened. This principle is based on the fact that the primary aim of an orthography is to convey meaning (Stark, 2010). The underrepresentation caused by the use of <b> for both /^{m}b/ and /β/ creates ambiguity and diminishes differentiation to both the writer and the reader of Gĩkũyũ. Two words with different meanings, for example <**mb**athi> (bus) and <**b**athi> (iron box) will be seen to have the same meaning (see (1 a) above). Disambiguation will demand greater effort for the reader. To cater for this, we propose the constraint CM- (MEAN)

Nevertheless, the graphemes adhere to P3 which requires that speakers write as they speak. The graphemes represent the physical realizations that some speakers produce; that is, [b], [d] and [g]. Thus, constraint IDENT- (F, P) becomes relevant in this case.

P5, which militates against changing the graphemic inventory of a language, is also flouted. For example, the two graphemes <d> and <j> do not exist in the graphemic inventory of Gĩkũyũ and so should not be in use. This necessitates weighing of the graphemes against the markedness constraints INV-(G).

In Gĩkũyũ, graphemes for prenasalised stops such as <mb>, <nd>, <ng> and <nj> may, and do occur, in the word initial position. However, in English, such combinations of letters, never occur in this position. The voiced stops in that position are represented by <b>, <d>, <g> and <j>. We theorize that due to this fact, some Gĩkũyũ speakers overlook the part of the grapheme that represents the nasal segment in the word initial position and simply omit it. Using the graphemes used in English observes P6 but creates a conflict with the principles under the phonological adequacy criterion.

In addition, the use of digraphs in the representation of the nasal segment in /^{m}b/, / ^{n}d/, / ^{ŋ}g/ and /^{nj}ɟ/ is tedious as expressed by some speakers in the study. The resultant graphemes interfere with the visual salience of the speakers as they are neither easy to write nor read. As a result, the writers choose to disregard or simply ignore the digraphs and thereby flout P7. The relevant constraint in this case is *COMP.

Many of the respondents see <ng>, <mb>, <nd> and <nj> as overlapping with <g>, <b>, <d> and <j>, respectively, which are graphemes found in English. They thus use the graphemes used in English and so contravene P8. The graphemes <mb> and <b> are a source of confusion to some of the users because they call for conscious effort in making a choice.

The constraints for the plosives and the affricate are ranked as shown below:

INV-(G) >> CM-(MEAN), NC-(N, P) >> *COMP>>MAX-(1P, 1G), IDENT-(1G, 1P), DEP-(P, G), IDENT-(FP)

4.5.1.1 The Voiced Bilabial Plosive /^{m}b/

In the analysis of /^{m}b/, both candidates <mb> and <b> appear in the graphemic inventory of the language and so, the INVENTORY constraints are not in consideration.

Tableau 1: Representation of /^{m}b/

/^{m}b/<mb>	NC- (N, P)	*COMP	MAX- (1P,1G)	ID- (1G,1P)
a. ☞mb		*		
b. b	*!		*	*

Grapheme (b) violates MAX- (1P, 1G) because it is not the specified grapheme for the phoneme, and violates ID- (1G, 1P) because it represents another phoneme in Gĩkũyũ. Nevertheless, these are low ranked constraints. Though it does not violate *COMPLEX in that it is made up of just one symbol, it however, violates NC-(N, P) because it does not represent the pre-nasalised voiced plosive /^{m}b/. This is a fatal violation because it goes against the phonotactics of Gĩkũyũ.

Candidate (a) is the optimal candidate. It does not violate the low ranked faithfulness constraints. However, it violates *COMPLEX because it is composed

of two symbols. Nevertheless, it does not violate the higher ranked NC - (N, P) and so is the winner.

4.5.1.2 The Alveolar Plosive /^{n}d/, <nd>

There are two candidates for analysis, <nd> and <d>. The second candidate is not in the graphemic inventory of Gĩkũyũ. Therefore, we consider the INVENTORY constraints. We also consider the NC-(N, P) constraints that require that plosive representation should cater for pre-nasalisation. The *COMP constraints are presented because one of the candidates is complex in appearance. The input-output relations are taken care of by the MAX - (1P, 1G) constraints.

Tableau 2. Representation of /^{n}d/

/^{n}d/ <nd>	INV-(G)	NC-(N, P)	*COMP	MAX- (1P, 1G)
a.☞nd			*	
b. d	!*	*		*

Candidate (b) violates MAX- (1P, 1G) because the input does not match the output for /^{n}d/. It also violates NC - (N, P) since it leaves out representation of the pre-nasal segment. It fatally violates INV-(G) since it is not in the graphemic inventory of Gĩkũyũ.

Candidate (a) violates *COMP since it is made up of two segments. However, it shows the representation of the pre-nasal segment and so does not violate NC-(N, P). It is also a grapheme in the inventory of Gĩkũyũ and so wins by not violating this highest ranked constraint.

4.5.1.3 The Velar Plosive /ᵑg/, <ng>

There are two candidates generated for this phoneme, that is <ng> and <g>. We pit them against the markedness constraints of NC-(N, P) that require the representation of the prenasal segment, and *COMP that require simplicity of graphemes. The faithfulness constraints that apply here are MAX-(1P, 1G) that require a one phoneme, one grapheme correspondence, IDENT-(1G,1P) that require a one grapheme, one phoneme correspondence, and IDENT-(F, P) that calls for correspondence in features.

Tableau 3 that follows shows the analysis.

Tableau 3. Representation of /ᵑg/

/ᵑg/<ng>	NC- (N, P)	*COMP	MAX- (1P, 1G)	IDENT- (1G, 1P)	IDENT- (F, P)
a. ☞ng		*			
b. g	!*		*	*	*

Candidate (b) is a grapheme that represents another phoneme. This is a violation of IDENT-(1G, 1P), and of IDENT-(F, P) because there is no correspondence in features. It is also not the representation of the given input which violates MAX-(1P, 1G). Though it is easy to write and, therefore, does not violate *COMP, it does not show representation of the prenasal segment and so violates NC-(N, P). This makes it disharmonious and eliminates it.

Candidate (a) is the most optimal candidate. It does not violate any of the faithfulness constraints. However, it is made up of two segments which renders it complex, and so it violates *COMP. Nevertheless, it does not violate the other higher ranked constraints in that it conveys the correct meaning and also represents the prenasal segment that is present in the phoneme.

4.5.1.4 The Palatal Affricate /ᵐɟ/, <nj>

There are two candidates, <nj> and <j>. The markedness constraints against which the candidates are weighed are INV-(G) that requires the grapheme be in the graphemic inventory of Gĩkũyũ, and *COMP that requires simplicity of graphemes. The faithfulness constraints at play here are MAX-(1P, 1G) and IDENT-(FP)

Tableau 4: Representation of /ᵑɟ/

/ᵑɟ/<nj>	NC- (N, P)	*COMP	MAX- (1P, 1G)	IDENT- (FP)
a. ☞nj		*		
b. j	!*		*	*

Candidate (b) violates IDENT-(FP) since it represents a palatal approximant while the input is a prenasalised affricate. It also violates MAX-(1P, 1G) because it is not the grapheme for the input. Though it does not violate *COMP in that it is simple in constitution, it nevertheless, violates NC-(N, P) because it does not adhere to Gĩkũyũ graphemic structure where an affricate must be prenasalised. This is a fatal violation which eliminates it.

Candidate (a) is the winner. It only violates the *COMP constraints because it is made up of two segments but observes the higher ranked INV-(G). This makes it the more harmonious of the two.

4.5.2 The Nasals

There are four Gĩkũyũ nasals, /m/, /n/, /ŋ/ and /ɲ/ graphemically represented as <m>, <n>, <ng'>, and <ny>, respectively.

In our data, both the bilabial nasal /m/ and the alveolar nasal /n/ had one surface realization each, [m] and [n], respectively. Each and a singular graphemic representation as designed in the inventory. The observation here is that there was adherence to the orthography principles in setting up the orthography and that the speakers keep to them**.**

The velar nasal /ŋ/ though, has only one surface realization [ŋ], but there are two graphemic representations, <ng'> and <ng>. The palatal nasal /ɲ/ also records one phonemic realization [ɲ], and three graphemic representations, <ny>, <nj> and <y>. For both phonemes, P1 and P2 are flouted. There is neither One phoneme-One-grapheme nor One grapheme-One-phoneme correspondence. This is a violation of MAX-(1P, 1G), MAX-(1G, 1P), IDENT-(1G, 1P) and DEP-(P, G).

Principle P3 is also contravened since only one phone per phoneme is articulated yet there are many representative graphemes. ID-(FT) constraints that require one to write as they speak, are considered here. Graphemes <ng>, <nj> and <y> contravene P4, the Convey Meaning Principle. The graphemes are also used to represent /ᵑg/, /ᶮɟ/ and /j/, respectively. Their use creates underrepresentation (see §5.2). This in turn leads to ambiguity. We, therefore, consider the markedness constraints of CM-(MEAN).

The grapheme in the Gĩkũyũ graphemic inventory for the velar nasal <ng'> contravenes both P6 and P7. Some speakers view the grapheme as cumbersome and confusing to write because of the diacritic. In English, the diacritic (') occurs

as an apostrophe to indicate possession or omission of a letter or letters. This is a source of confusion to some speakers who must use the diacritic to indicate a new phoneme. This contravenes P6. Some speakers avoid using it, thus changing the grapheme to <ng> that represents /ᵑg/, and which has a heavier functional load. Though this change takes place, speakers still read the given word accurately in its totality either using the context or their phonological knowledge of the language. An example is the word '*Ng'ang'a'* (name of a person) which speakers decipher accurately and separately from '*nganga'* (a guinea-fowl). This flouts P7. The relevant universal markedness constraints emanating from these two principles are *DIA and *COMP.

In regard to P9, the nasal graphemes are all accepted except for <ng'>. This grapheme does not represent /ŋ/ in English but it does so in Kiswahili, for instance in the words **ng'**oa, 'uproot' and **ng'**ambo, 'abroad'. The respondents who did not use it were from Ndia, Gĩcũgũ and the Northern dialects. They write it as <ng>, the grapheme that represents the same sound, /ŋ/, in English. They indicated that they found it cumbersome to keep on representing the sound with the grapheme specifically because of the apostrophe. This implies a rejection of the grapheme based on the premises that it is cumbersome to write. In using grapheme <ng>, P9 is flouted. This calls for the faithfulness constraint, ACCEPT.

The ranking of the constraints is as shown below.

CM-(MEAN) >> *COMP, *DIA >> MAX-(1P,1G), MAX-(1G,1P), IDENT-(1G,1P), DEP-(P, G), IDENT-(FT), ACCEPT

The analysis using tableau for the two nasals is given below.

4.5.2.1 The Palatal Nasal /ɲ/ Grapheme <ny>

There are two candidates, <ny> and <nj>. These are pitted against the MAX- (1G, 1P), IDENT-(1G, 1P) and DEP-(P, G) constraints. MAX-(1P, 1G) constraints are not considered because none of the candidates deletes a segment in its representation. The candidates are not pitted against *COMP constraints because they are both made up of more than one symbol, and not against *DIA because neither has a diacritic. CM-(MEAN) is the determining constraint for them.

Tableau 5: Representation of /ɲ/

/ɲ/ <ny>	CM-(MEAN)	MAX-(1G, 1P)	IDENT-(1G, 1P)	DEP (P,G)
a. ☞ny				
b. nj	!*	*	*	*

Candidate (b) is the loser. It violates all the constraints. It has no correspondence with the input which violates DEP-(P, G). It introduces a new component, the voiced prenasalised affricate, /ᵑɟ/ and so violates IDENT-(1G, 1P). It is also not the

representation of /ɲ/, which makes it violate MAX-(1G, 1P). These are lowly ranked constraints. Crucially, it violates CM-(MEAN) because it carries the meaning of another phoneme. An example would be the words **ny**a**ny**a (tomato) contrasted with **ny**a**nj**a (big calabash). This is a fatal violation.

Candidate (a) proves to be the optional candidate because it does not violate any of the constraints.

4.5.2.2 The Velar Nasal /ŋ/ <ng'>

There are two candidates, <ng'> and <ng>. In the analysis, *COMP constraints are not considered because both candidates are complex in their constitution. *DIA constraints that mitigate for ease in writing are considered. Also considered are VIS- (F) constraints that mitigate for the representation of salient features. In this case, the salient feature is nasality at the velar position. The faithfulness constraints that are considered are MAX- (1P, 1G), MAX-(1G, 1P) and DEP-(P, G).

Tableau 6: Representation of /ŋ/

/ŋ/ <ng'>	VIS-(F)	*DIA	MAX- (1P, 1G)	MAX- (1G, 1P)	DEP- (P, G)
a. ☞ng'		*			
b. ng	!*		*	*	*

Candidate (b) violates all the lower ranked faithfulness constraints. There is no correspondence between it as a representation for a prenasalised voiced velar plosive and the input which is a velar nasal. This flouts DEP-(P, G). The candidate is the grapheme for another phoneme and this violates MAX-(1G, 1P). It is also not the grapheme for the input which violates MAX-(1P, 1G). However, the three constraints are not decisive in the identification of the more harmonic candidate. The candidate observes the higher ranked *DIA-(F) because, between the two, it is the easier one to write; the other one has a diacritic. However, it fatally violates VIS-(F) because it is not the representation of the velar nasal.

Candidate (a) is the winner. It does not violate the low ranked constraints but violates *DIA-(F). However, this is not decisive because though diacritics make writing difficult, they may, nevertheless, carry prominent features.This is catered for by VIS-(F) constraints which are ranked higher. Since the candidate does not violate these constraints, it emerges as the more harmonious of the two.

4.5.3 The Fricatives

There are five Gĩkũyũ fricatives: the voiced bilabial /β/, the voiced interdental /ð/, the voiceless palatal /ʃ/, the voiced velar /ɣ/, and the voiceless glottal /h/. Each one of them is analysed separately in the sections below.

4.5.3.1 The Voiced Bilabial Fricative /β/

The voiced bilabial fricative /β/, had four surface realizations, [β], [ɸ], [v] and [f], and five graphemic representations, <b>, <mb>, <p>, <v>, and <f> (see Table 4.19).

The representation using the various graphemes contravenes P1, the One phoneme-One grapheme Principle since one phoneme is being represented by many graphemes. This lowers the speaker's ability to predict which of the graphemes should be used, and therefore, interferes with accuracy. It renders it problematic for the reader to determine which phoneme is represented unless the context is given. The use of the many graphemes for the phoneme constitutes overrepresentation (see §5.3). The constraints relevant to this principle are MAX-(1P, 1G).

P2, the consistency principle, is also contravened. The use of grapheme <mb> violates the principle which militates against using one grapheme to represent more than one phoneme. This grapheme is used to represent phoneme /^{m}b/, thus invoking MAX-(1G, 1P), IDENT-(1G, 1P) and DEP-(P, G) constraints.

P3, the Write as You Speak Principle, is observed. The speakers physically produce [β], [ɸ], [v] and [f]. In using the different graphemes, they are representing each of the phones. Though this principle is observed, the outcome is in conflict with P1 and P2. The constraints that are at play here are IDENT-(FT) that mitigate for representation of phones.

The graphemes <p>, <v>, and <f> are not in the inventory of Gĩkũyũ graphemes. Their usage violates principle P5 that disallows changing the inventory of letters. This invokes the markedness constraints of INV-(G).

The use of <b> for /β/ is a contravention of P6. In both English and Swahili, <b> represents /b/, but in Gĩkũyũ, it represents /β/. Both /b/ and /β/ are voiced bilabials but they do not share the manner of articulation, with /b/ being a plosive and /β/ a fricative. An even closer phoneme to /β/ is /ɸ/ which differs only in voice. The phoneme is found in Kikamba (Kioko, Njoroge & Kuria, 2012) and is graphemically represented with <v>. Kimbeere and Kiembu languages have the labio-dental fricative /v/ which is represented with <v> (Iribemwangi, 2012). These three are Bantu language that are closely related to Gĩkũyũ. The Gĩkũyũ speakers are in close geographical proximity to speakers of these languages and there is high social interaction. Speakers using grapheme <v> in Gĩkũyũ probably deem it appropriate to use a grapheme that represents a similar phoneme in a closely related language. A constraint that can be invoked here is the faithfulness constraint IDENT-(F, U) which requires that familiar symbols should not be used in unfamiliar ways. We also invoke the markedness constraint *COMP because one of the candidates is made up of more than one segment. The constraints can be ranked thus:

> INV-(G) >> *COMP >> MAX-(1P, 1G), MAX-(1G, 1P), IDENT-(1G, 1P), DEP-(P, G), IDENT-(F, U)

Constraints MAX-(1G, 1P), IDENT-(1P, 1G) and IDENT-(F, U) are violated by all but one candidate. Since they all rank low, they are not considered in the analysis. The five candidates are analysed as shown in Tableau 7 below.

Tableau 7: Representation of /β/

/β/ <b>	INV-(G)	*COMP	MAX- (1P, 1G)	MAX- (1G, 1P)	DEP- (P, G)	IDENT- (F, U)
a.☞b						
b. mb		!*	*	*		*
c. p	!*		*			*
d. v	!*				*	
e. f	!*		*		*	

Candidate (f) violates DEP-(P, G) by inserting the feature [+ dental] which is not present in the input. It violates MAX-(1P, 1G) by deleting the feature [+voice] that is present in the input. It loses by fatally violating INV-(G) because it is not in the inventory of the language. Candidate (d) also inserts the feature [+dental] thereby violating DEP- (P, G) and is eliminated for violating INV-(G). Candidate (c) violates IDENT-(F, U) because no respondent produced the voiceless bilabial plosive /p/. The deletion of the [+voice] feature makes it violate MAX-(1P, 1G),

but it is the violation of INV-(G) which proves fatal to it. Candidate (b) violates the lowest ranked constraint. In place of the given input, respondents do not produce the phone that the candidate represents. The candidate also violates *COMP which knocks it out of contention. Candidate (a) proves the most harmonic, and so the winner by not violating any of the constraints.

4.5.3.2 The Voiced Inter-Dental Fricative /ð/

The voiced inter-dental fricative /ð/ has three surface realizations, [ð], [^{n}d] and [z] and four graphemic realizations, <th>, <nd>, <d> and <ny>. The use of the many graphemes contravenes P1, the One phoneme - One grapheme Principle. The MAX-(1P, 1G) constraints are, therefore, relevant. As seen in (4b) above, a reader may find it difficult to identify the grapheme to use accurately. The consistency principle, P2, is also violated because <nd> and <ny> represent other phonemes, the pre-nasalised voiced alveolar plosive /^{n}d/ and the palatal nasal /ɲ/. Constraints MAX-(1G, 1P), IDENT-(1G, 1P) and DEP-(P, G) apply in this case. There is no knowing which grapheme to write in whichever word position. P3, the Write as You Speak Principle, is flouted since what are being represented are not the phones that are even associated with /ð/. This invokes IDENT-(FT) constraints. P4 is violated because <nd> and <ny> represent other phonemes, /^{n}d/ and /ɲ/, respectively. This conveys different meaning and so CM-(MEAN) constraints become applicable.

Grapheme <d> is in violation of P5 as it is not in the inventory of Gĩkũyũ graphemes.The INV-(G) constraints are therefore in consideration. The graphemic representations for /ð/ present contraventions to the simplicity criterion. The distinction between <t> and <th> presents problems to some speakers of the Gĩcũgũ dialect. The 'h' in <th> is seen as representing a phoneme which is not found in the dialect. This makes writing or reading the grapheme difficult to them. This flouts P6 and violates *COMP.

The ranking of the constraints is as shown below.

> INV-(G) >> CM-(MEAN), *COMP>> MAX-(1P, 1G), MAX-(1G, 1P), IDENT-(1G, 1P), DEP-(P, G), IDENT-(F, U).

MAX-(1P, 1G) will not be considered because all candidates except one, <th>, violate it by deleting the fricative element, and it is not a decisive constraint. Likewise, all the candidates except <th> violate the lowly ranked DEP-(1G, 1P) by hampering the consistency of one grapheme one phoneme, and so the constraint will not be considered. *COMP constraints are considered because some candidates are made up of more than one segment. CM-(MEAN) are also considered because some candidates convey meanings that differ from the one intended by the input. The analysis is presented in the tableau below.

Tableau 8: Representation of /ð/

/ð/<th>	INV-(G)	CM-(MEAN)	*COMP	IDENT-(1G,1P)	DEP-(P, G)	IDENT-(F, U)
a.☞th			*			
b. nd		!*	*	*	*	
c. d	!*	*		*	*	
d. ny		!*	*	*	*	*

Candidate (d) violates IDENT-(F,U) because it does not represent the phone for /ð/. It violates DEP-(P, G) by inserting nasality into the fricative. It also violates IDENT-(1G, 1P) because it represents another phoneme, /ɲ/. Its violation of CM-(MEAN) by carrying the unintended meaning eliminates it. Candidate (c) does not violate the lowest ranked constraint because some speakers produce a phone nearest to this, a dentalised voiced alveolar plosive /d_n/. However, it inserts a [+stop] element to replace the fricative element and is not the grapheme for the phoneme. Though it carries no meaning and so violates CM-(MEAN), what eliminates it is its violation of INV-(G) because it is not a grapheme in Gĩkũyũ. Candidate (b) inserts representation of an alveolar thus violating DEP-(P, G), and also violates IDENT-(1G, 1P) because it represents another phoneme, /^{n}d/. The

meaning it conveys is not the intended one and so it violates CM-(MEAN). This eliminates it. Candidate (a) is the most optimal. It only violates *COMP constraints but carries the meaning intended by the input.

4.5.3.3 The Voiceless Palatal Fricative /ʃ/

The voiceless palatal fricative /ʃ/ had three surface realizations [ʃ], [tʃ] and [s] and the graphemes presented by the respondents were three, <c>, <s> and <ch>. The use of the three graphemes for one phoneme contravenes P1. Both <ch> and <s> are used by speakers who physically realize the phoneme as [tʃ] and [s] and so P3 is observed. On the other hand, <ch> and <s> are not in the graphemic inventory of Gĩkũyũ and so P5 is flouted. This calls for INV-(G) constraints.

The use of <c> flouts P6. The grapheme is used to represent /ʃ/ in Gĩkũyũ. In Kiswahili the grapheme is not in use but in combination with <h> (written as <ch>), it represents /tʃ/. In English <c> represents three phonemes which are /k/ (in **c**ake), /s/ (**c**ite) and /ʃ/ (musi**c**ian). The decision to use the grapheme is thus a contravention of this principle. A more appropriate grapheme would be <s> which is familiar in both English and Swahili, and which would at least represent the sound produced by some Gĩkũyũ speakers. One of the constraints invoked here is *COMP. Another one is ACCEPT while the final one is IDENT-(F, U) that requires that familiar symbols should not be used in unfamiliar ways.

Some respondents feel /ʃ/ should not be represented by <c> but rather by <s> or <sh>. Some feel that /β/ should be represented by <f> and not <b>. This indicates

that both <c> and <b> are not wholly acceptable to the speakers which is a contravention of P9. The ranking of the constraints is as follows:

INV-(G) >> *COMP >>IDENT-(F, U), IDENT-(FT)

The analysis of the graphemes is shown in the tableau below.

Tableau 9: Representation of /ʃ/

/ʃ/ <c>	INV-(G)	*COMP	IDENT-(F, U)	IDENT-(FT)
a.☞c			*	*
b. s	!*			
d. ch	!*	*		

Candidate (d) observes IDENT-(FT) because some speakers produce [tʃ]. It also observes IDENT-(F, U) because some speakers use a familiar symbol though in an unfamiliar way. In addition, it is complex and so violates *COMP. However, it is the violation of INV-(G) that renders it disharmonic and so it loses out.

Candidate (b) does not win either. Though it violates none of the faithfulness constraints as well as *COMP, it nevertheless, violates INV-(G) which is the highest ranked and decisive constraint.

Candidate (a) is the winner. It violates IDENT-(FT) because it appears like the representation of the voiceless palatal plosive [c]. It also violates IDENT-(F, T) in that it represents a familiar grapheme used in an unfamiliar way. It, however, does not violate the decisive INV-(G) because it is in the graphemic inventory of the Gĩkũyũ. Thus, it is the most harmonic.

4.5.3.4 The Voiced Velar Fricative /ɣ/

The voiced velar fricative /ɣ/ had two surface realizations, [ɣ] and [k]. There were two graphemic representations to it, <g> and <k>. The use of two graphemes contravenes P1, the One phoneme-One grapheme Principle. P2, the One grapheme-One phoneme Principle, is contravened because <k> represents another phoneme, the voiceless velar plosive /k/. Since there are speakers who produce [k], P3, the Write as You Speak Principle is observed. Grapheme <k> represents /k/ and so its usage contravenes P4, the Convey Meaning Principle. Its use causes confusion in that different lexical items are assigned the same meaning, for example, 'ri**g**a'(be unknown) and 'ri**k**a' (dive into/ dip/ settle). We consider the markedness constraints of CM-(MEAN) because of this. The ranking of the constraints is:

CM-(MEAN) >> MAX- (1P, 1G), IDENT-(1G, 1P), DEP- (P, G)

The analysis of the two candidates is presented in the following tableau.

Tableau 10: Representation of /ɣ/

/ɣ/ <g>	CM- (MEAN)	MAX- (1P, 1G)	IDENT- (1G, 1P)	DEP- (P, G)
a.☞g				
b. k	!*	!*	*	*

Candidate (b) violates DEP-(P, G) by inserting a new feature, [+plosive]. It represents the grapheme for another phoneme and so violates IDENT-(1G, 1P). The candidate also deletes the features [+fricative, +voice] from the input and so violates MAX-(1P, 1G). The violation of CM-(MEAN) knocks it out of contention against the other candidate. Candidate (a) is optimal. It wins because it violates none of the constraints under consideration.

4.5.3.5 The Glottal Fricative /h/

The glottal fricative /h/ had two surface realizations, [h] and [b]. The graphemes presented were <h>, <b> and <mb>. The use of many graphemes means that there is no one phoneme one grapheme correspondence, a violation of P1 and the MAX-(1P, 1G) constraint. In using <b> and <mb>, the respondents contravene P2, the One grapheme -One phoneme Principle since these graphemes also represent other phonemes. Grapheme <b> represents the voiced bilabial fricative /β/ and <mb> represents the pre-nasalised bilabial plosive <^{m}b>. P3 is not contravened

since the graphemes represent the phones that are produced. Graphemes <b> and <mb> also represent other phonemes and their use interferes with meaning. This flouts P4 and violates CM-(MEAN).

The Ndia and Gĩcũgũ dialects do not have the phoneme /h/ in their inventory (Iribemwangi, 2012). In its place, the two produce /β/ and /^{m}b/ in both word initial and medial positions. This is fronting and consonant switching which can compare with Japanese. It is observed that in Japanese, the phoneme /h/ is realized as [ɸ] when it occurs before /u/ (Ruddell, 2013). In Gĩcũgũ, phoneme /h/ switches to either [β] or [^{m}b]. The use of both <b> (**b**uu**b**u) and <mb> (**mb**ubu) instead of <h> (**h**uu**h**u, "bats") indicates that some speakers of the two dialects do not accept <h> as representative of the phoneme /h/. This flouts P9, the Principle on Acceptability. Consequently, ACCEPT constraints are invoked. The constraints can be ranked as follows:

CM-(MEAN) >> ACCEPT, IDENT-(1G, 1P), MAX-(1P, 1G)

The analysis is presented in the tableau below.

Tableau 11: Representation of /h/

/h/ <h>	CM- (MEAN)	ACCEPT	IDENT- (1G, 1P)	MAX- (1P,1G)
a.☞h		*		
b. b	!*		*	*
c. mb	!*		*	*

Candidates (c) and (b) do not represent the designated phoneme and, therefore, violate MAX-(1P, 1G). Additionally, both represent another phoneme, /^{m}b/ and /β/, respectively. This violates IDENT-(1G, 1P). The fact that these candidates convey the unintended meaning makes them fatally violate CM-(MEAN). The most optimal candidate is (a). It violates the low ranked ACCEPT but not the higher ranked one, thus emerging the winner.

4.5.4 The Alveolar Tap /ɾ/

The tap had one surface realization [ɾ] but two graphemic representations, <r> and <l>. The grapheme <l> represents the lateral /l/. Both [ɾ] and [l] are liquids and, though articulated differently, are acoustically similar (Fromkin et al., 2007). It could be due to this similarity that some respondents substitute <l> for <r>.

The representation with <l> interferes with accuracy because one phoneme is being represented by more than one grapheme. This flouts P1. Inconsistency is created since it is unclear what grapheme to use in the representation of the phoneme. Using <l> observes P3, the Write as You Speak principle. As noted above, some respondents perceive /l/ in the production of /ɾ/ and so represent their own perception. However, this conflicts with P1.

Since <l> does not exist in the graphemic inventory of Gĩkũyũ, P5 is contravened and so, INV-(G) constraints are considered. In addition, the meaning communicated in a word that has the grapheme requires more effort to decipher since it is a strange symbol. This flouts P4.

The grapheme <l> was used in the initial word position by some respondents from all the Gĩkũyũ dialects. It should be noted here that the word carrying the phoneme (**r**ita) is a borrowed word from English (litre). The use of the grapheme is clearly a case of the influence of the lender language. This is a case of using a familiar symbol in an unfamiliar way and is therefore a contravention of P6. However, this is despite the fact that respondents who used it had indicated that writing in Gĩkũyũ was easy and so it is expected that they would not have used the grapheme.

The respondents in this study did not indicate that the tap was confusing. This gives the implication that they are comfortable using the grapheme <r> and so accept it, which is an observance of P8. Nonetheless, when it comes to writing,

some respondents across the dialects (except the Southern) use grapheme <l> for the tap. Additionally, a male from Gĩcũgũ and a female from Ndia suggested that <l> should be removed from the Gĩkũyũ graphemic inventory, while, in fact, it is not in the inventory. From this, we posit that Gĩkũyũ speakers are aware of grapheme <l>, but due to the closeness between [ɾ] and [l], they switch the consonant representation and end up producing the two graphemes. This violates P9.

The two candidates, <r> and <l> are analysed against INV-(G), MAX-(1P, 1G) and ACCEPT constraints.

Tableau 12: Representation of /ɾ/

/ɾ/ <r>	INV - (G)	MAX- (1P, 1G)	ACCEPT
a. ☞r			
b. l	!*	*	*

Candidate (b) is not accepted by some respondents and so violates ACCEPT. Since it interferes with the One phoneme-One grapheme accuracy, it also violates MAX-(1P, 1G). However, its violation of INV-(G) is fatal and eliminates it.

Candidate (a) wins. It does not violate any of the constraints and so is the optimal candidate.

4.5.5 The Front High [+ATR] Vowel /i/

Turning to vowels, Githiora (2003, p. 86) identifies <ĩ> and <ũ> as vocalic graphemes that were "bones of contention" during the orthography design. Indeed, Wanjau (1989) as quoted in Kuria (2005) observes that some discrepancies with the vocalic phonemes were resolved by the United Kikuyu Language Committee. From our data, we contend that many discrepancies are still found between the phonemes and their graphemes.

The front high [+ATR] short vowel, /i/ is presented by all the respondents phonetically as [i] and phonemically as /i/. Two graphemes are used to represent the phoneme. These are <i> and <ĩ> in all the word positions; that is, initial, medial and final positions.

The use of the two graphemes contravenes P1 because one phoneme is being represented with various graphemes. The two graphemes are used in Gĩkũyũ to represent /i/ and /e/. Phoneme /e/ shares the phonetic characteristic [+ATR] with /i/ and is only different because it is lower and therefore more centralized. The use of the two graphemes infringes on maximization of contrast that engenders maximization of perceptual salience (Boersma, 1998). Accuracy is reduced and overrepresentation is created. In this case, we consider the MAX-(1P, 1G) constraints.

The two graphemes also flout Principle 2 on consistency. Only one grapheme should be used for the phoneme. The inconsistency creates unpredictability in deciding whether to use <i> or <ĩ>. We therefore, consider the MAX-(1G, 1P) constraints.

Principle 3 is contravened because, though the respondents articulate [i], they use <ĩ> which represents what they do not articulate. This brings in the constraint IDENT-(FP).

The use of the two < i > and < ĩ > contravenes P4 which deals with conveyance of meaning. Meaning is lost in trying to decipher exactly which phoneme to use, and therefore, which meaning is being expressed. This is underrepresentation and it causes ambiguity as evidenced between the words <**i**ta> 'strangle' and < **ĩ**ta > 'call'. As already noted, this interferes with the principle of maximization of differentiation (Boersma, 2014). The relevant markedness constraint here is CM-(MEAN).

The ranking of the constraints is as given below:

CM-(MEAN) >> IDENT-(1G,1P), IDENT-(FP)

The analysis of /i/ is presented in the following tableau.

Tableau 13: Representation of /i/

/i/ <i>	CM-(MEAN)	IDENT-(1G,1P)	IDENT-(FP)
a. ☞i			
b. ĩ	!*	*	*

Candidate (b) violates the faithfulness constraints. It loses out because it violates CM-(MEAN) whereby, it conveys an unintended meaning. Candidate (a) wins by not violating any of the constraints.

The representation of the long vowel /i:/ also flouts some principles in the adequacy criterion. In the word initial position, the grapheme used by the respondents is either <i> or <ii>. Doubling of the grapheme in this position should not occur due to predictability (Armstrong, 1967). Other supra-segments such as tone may be at play and may overshadow vowel length but doubling is disallowed. Additionally, the orthography convention is that there should be no doubling in this position (Githiora, 2003; Stegen, 2005). Doubling is a flouting of P1 because there is no one phoneme to one grapheme correspondence. P2 is also flouted since <i> represents /i/ and /i:/ and <ii> represents /i:/. On the other hand, tone and stress instead of vowel length may have been perceived by the respondents and so P3 may have been observed. The intended meaning is

conveyed and so P4 is not flouted. The markedness constraint most applicable here is *VIS-(V-L) which mitigates against representation of vowel length. The evaluation is presented in the tableau below.

Tableau 14: Representation of /i:/ in word initial position

/i:/ <ii>	*VIS (V-L)	MAX (1P, 1G)	MAX (1G,1P)	IDENT (FP)
a. ☞i			*	*
b. ii	*!	*		

Candidate (a) is the winner. Though it violates IDENT-(FP) because it does represent the feature [+length] that is in the input, it does not violate *VIS-(V-L) which militates against the representation of length. On the contrary, candidate (b) violates IDENT-(1G, 1P) by not being the grapheme for the phoneme, and also violates MAX-(1P, 1G) because of hampering consistency. The candidate fatally violates *VIS-(V-L) because it overtly represents vocalic length. This, therefore, eliminates it.

In the word medial and final positions, the graphemes used to represent length are the same as in the word initial position, <i> and <ii>. Both P1, the One phoneme-One grapheme Principle, and P2, the Consistency Principle are flouted. As is the

case with representation in the initial position, P3 is not flouted since some speakers may not have perceived vowel length but rather, tone and stress and so write as they speak. P4 is observed since the intended meaning remains the same. The markedness constraints in the analysis of the phoneme are inverted so that VIS-(V-L) now ranks higher and is, in fact, decisive. The *COMP constraint is also under consideration.

Tableau 15: Representation of /i:/ in Word Medial and Final positions

/i:/ <ii>	VIS (V-L)	*COMP	MAX (1P, 1G)	IDENT (1G,1P)	IDENT (FP)
a. ☞ii		*			
b.i	!*		*	*	*

Candidate (b) is the loser. It violates all the faithfulness constraints. By not overtly representing vowel length, it fatally violates the critical and decisive VIS-(V-L). Candidate (a) is the winner and the optimal candidate because it does not violate any of the low ranked faithfulness constraints and only violates *COMP because it is made up of more than one symbol. It emerges the winner by not violating the highest ranked constraint, VIS-(V-L).

When /i/ is concatenated with /u/ and /ɔ/ to give graphemes <iu> and <io>, respectively, no discrepancies occur. It is when /i/ is concatenated with /e/, /o/ and /a/ so as to be represented as <iĩ>, <iũ> and <ia> respectively, that discrepancies occur. For /ie/, graphemes <iĩ> and <ii> are presented. For /io/, the graphemes presented are <iu>, <ĩũ>, <io> and <ĩu> in addition to <iũ>. For /ia/, grapheme <ĩa> is also presented. These graphemes contravene P1, P2 and P3. The Convey Meaning Principle, P4 is nevertheless, maintained since the speakers express, through reading, what is intended. P5 is not contravened because the graphemes that have been used are in the inventory of Gĩkũyũ.

The graphemes <i> and <e> for the phonemes /i/ and /ɛ/ respectively, were confusing to some of the respondents. The respondents, being familiar with the phonology of English and the way these graphemes are used in English, face some difficulties. In English, <e> represents /e/ in ***e****dge*, /ɪ/ and /i/ in *d****e****l****e****te*, /ɛ/ ***e****lephant* and /a/ in ***e****ye*. Grapheme <i> represents /ɪ/ in *p****i****t*, /ɜ:/ in *b****i****rd* and /ai/ in *wr****i****te*. Yet, in Gĩkũyũ, <i> represents /i/ only and <e> represents /ɛ/ only. This demands more conscious effort from both the writer and the speaker in making a choice that applies to Gĩkũyũ and not to English. Greater effort is demanded when one of the graphemes should be doubled so as to indicate length. The consequence is that the writer ends up using either of the two and leaves it to context for the reader to decipher exactly what was desired. This conflicts with both P6 and P7. The constraints that are violated include the faithfulness IDENT-(F, P) and the CM-(MEAN).

Graphemes <i> and <e> do not create difficulties in writing nor is unacceptability expressed about them. With these two graphemes, P8 and P9 are observed.

4.5.6 The Mid [+ATR] Vowel /e/

Vowel /e/ is realized on the surface as [e]. There are two graphemes that are presented, the designated grapheme <ĩ> on the one hand and <i> which represents the high front [+ATR] /i/ on the other. The use of the two graphemes flouts P1 since accuracy is reduced. This violates MAX-(1P, 1G). Again, inconsistency is created since <i> is also used for /i/. This flouts P2 and violates MAX-(1G, 1P). P3 is flouted, and subsequently, IDENT-(FP) is violated by <i> because the speakers produce [e] and therefore, do not write as they speak. This could possibly be explained by a merger of /i/ and /e/. P4 is also flouted which is a violation of CM-(MEAN) since there is ambiguity that is created by the underrepresentation of the phoneme /e/ (see the example of ***i****ta* 'strangle' and ***ĩ****ta* 'call' given above).

Grapheme <ĩ> is not easy to type on an ordinary keyboard or even on a telephone handset, a matter that respondents reported. When it comes to hand-writing, the grapheme is difficult to write in a fast - flowing manner, more so because of the diacritic/macron. Respondents said they found it cumbersome, both in reading and writing. This invokes the markedness constraints *NO COMP and *DIA.

The grapheme is hard to master without some teaching, learning and practice. Some speakers/ writers therefore avoid writing it, preferring to use <i> instead.

This contravenes principle P8 on the feeling of correctness that arises from habitual usage. This violates the faithfulness constraint, ACCEPT.

The ranking of the constraints is as shown below.

CM-(MEAN),*DIA, *COMP >> ACCEPT, MAX-(1P, 1G), MAX-(1G, 1P), IDENT-(FP)

The analysis is presented in the following tableau.

Tableau 16: Representation of /e/

/e/<ĩ>	CM-(MEAN)	*DIA	*COMP	ACCEPT	MAX-(1P, 1G)	MAX-(1G, 1P)	IDENT-(FP)
a. ☞ĩ		*	*	*			
b. i	!*				*	*	*

Candidate (b) violates most of the faithfulness constraints except ACCEPT. This is because this is the grapheme that most speakers are familiar with. It is easy to write and read and also has no diacritic. Thus, the candidate does not violate both *COMP and *DIA. However, this candidate carries an unintended meaning and so violates CM-(MEAN), the highest ranked constraint. This makes it a loser. Candidate (a) is the winner. Though it is rejected by some candidates and so

violates ACCEPT, it is also not easy to write, which violates *COMP. In addition, it has a diacritic which makes it violate *DIA. Nevertheless, it carries the intended meaning and so does not violate CM-(MEAN). Naturally then, it is the optimal candidate.

The long vowel /e:/ is represented with <ɨ̂>, <i>, < ɨ̃ɨ̃> and <ii> in all word positions. The last two graphemes in the initial word position flout the convention that vowel length should not be indicated in the initial position. In the word medial and final positions, doubling the grapheme to indicate length is cumbersome and interferes with ease in writing.

The graphemes used reduce accuracy and therefore flout P1. They also create inconsistency and thereby flout P2. Graphemes <ɨ̂>, <i>, and <ii> contravene all the principles under this criterion. The constraints that may be considered in the analysis include MAX-(1P, 1G) from P1, IDENT-(1G, 1P) from P2, IDENT-(F, P) and *VIS-(V-L) from P3, CM-(MEAN) from P4, *DIA and IDENT-(FP) from P6, and lastly, ACCEPT from P9. Their ranking is as follows:

> *VIS-(V-L) >> CM-(MEAN), *DIA, *COMP >> ACCEPT, MAX 1P,1G), IDENT-(1G,1P), IDENT-(FP)

The analysis of the candidates is presented in the following tableau.

Tableau 17: Representation of /e:/ in word initial position.

/e:/ <ĩ>	*VIS- (V-L)	CM- (MEAN)	*DIA	*COMP	ACCEPT	MAX 1P,1G)	IDENT- (1G,1P)	IDEN T-(FP)
a. ☞ĩ			*	*	*			*
b. i		!*				*	*	*
c. ĩĩ	!*		*	*	*	*	*	
d. ii		!*				*	*	*

Candidate (d) does not have the features of the input and so violates IDENT-(FP). Besides, it is not the grapheme for the input and so hampers consistency, and is also not the grapheme for the given phoneme. Additionally, it does not convey the meaning and so fatally violates CM-(MEAN). Candidate (c) violates the lower ranked constraints except IDENT-(FP). It violates *COMP because it is not easy to write and also violates *DIA because it uses complex symbols, the diacritics. It fatally violates *VIS-(V-L) that militates against the presentation of vowel length in the word initial position. Candidate (b) also violates the lower ranked constraints but crucially violates CM-(MEAN). Candidate (a) is the winner. It violates IDENT-(FP) because it does not represent the phone and neither is it accepted by some of the respondents. Though it is complex and also uses a

diacritic, it conveys the intended meaning. It also does not overtly represent vowel length and so does not violate *VIS-(V-L). The candidate, therefore, is the optimal one.

In the word medial and final positions, there are five candidates, <ĩĩ>, <ĩ>, <ii>, <i> and <ĩi>. The markedness constraints in this case are VIS-(V-L) which ranks highest, followed by CM-(MEAN). *DIA constraints come next followed, lastly, by *COMP. The faithfulness constraints emanate from P1, P2, and P3. MAX-(1P, 1G) and IDENT-(1G, 1P) are omitted from the analysis because four of the five candidates violate them.

The constraints are ranked as shown below:

VIS (V-L) >> CM (MEAN), *DIA, *COMP >> ACCEPT, IDENT-(FP)

The evaluation is as shown in the following tableau.

Tableau 18: Representation of /e:/ in word medial and final positions.

/e:/ <ĩĩ>	VIS (V-L)	CM (MEAN)	*DIA	*COMP	ACCEPT	IDENT-(FP)
a. ☞ĩĩ			*	*	*	
b. ĩ	!*		*	*	*	
c. ii		!*				
d. i		!*				*
e. ĩi			!*	*	*	

Candidate (e) does not violate IDENT-(FP) because some respondents actually produce the phoneme as represented by the candidate. The candidate is not accepted because it is not easy to write and has a diacritic. This combination knocks it out of contention. Candidate (d) violates the lowest constraint. It also conveys the unintended meaning which eliminates it. Candidate (c) only gets eliminated because it conveys a different meaning from the input. Candidate (b) is complex because of the diacritic but is only eliminated because it does not represent vowel length. Candidate (a) is, therefore, the winner. It fares much like candidate (b) with the only difference being that it represents vowel length, a feature that makes it the optimal candidate.

Concatenation of /e/ with other vocalic phonemes arouses discrepancies in all cases except when it is concatenated with the back high [+ATR] /u/. When concatenated with /i/ to be represented graphemically as <ĩi>, grapheme <ii> is additionally produced. The concatenation of the grapheme with <i> presents a problem. Writers usually present it simply as <i> without the diacritic. The reader is left to make out the intended phoneme from the context. P6 and P7 are thus flouted.

Concatenated with phoneme /o/ so as to be represented as <ĩũ>, graphemes <ĩu> and /iu/ are also produced. Concatenated with /ɛ/ to appear as <ĩe>, grapheme <ie> is additionally proffered, and with /ɔ/ to be represented with <ĩo>, <io> is given. Discrepancies are also noted when the phoneme is concatenated with /a/ to be represented with <ĩa> but <ia> is also generated.

The additional graphemes diminish accuracy and interfere with consistency in the writing system. This contravenes P1 and P2. The phones produced by the speakers in respect to the additional graphemes, do not tally with the phonemes under consideration. This is a contravention of P3. Meaning of words is lost because of the misrepresentations and so P4 is also contravened. Of note is that all the graphemes used are found in the inventory of Gĩkũyũ and so, P5 is not contravened.

4.5.7 The Front [-ATR] Vowel /ɛ/

The vowel was physically realized as [ɛ]. The grapheme for this phoneme is <e> which was produced along with <a>; thus, there are two candidates for analysis. The use of two graphemes hampers accuracy and is a contravention of P1. Inconsistency is observed since <a> also represents /a/ and /a:/. This contravenes P2. Grapheme <a> represents the low [-ATR] vowel. Its use here is indicative of a lowering of /ɛ/ to /a/ which flouts P4 by changing the intended meaning. Worth of note here is that the movement is from a [-ATR] vowel to another vowel that has the same tongue root value. The MAX-(1P, 1G), MAX-(1G, 1P), and IDENT-(FP), constraints that mitigate for correspondence in features, are invoked here. These are shown against the highest ranked markedness constraint of CM-(MEAN) in the tableau below.

The ranking of the constraints is:

CM-(MEAN) >> MAX- (1P, 1G), IDENT-(1G,1P), IDENT (FP)

Tableau 19: Representation of /**ɛ**/

/**ɛ**/ <**ɛ**>	CM- (MEAN)	MAX- (1P, 1G)	IDENT- (1G,1P)	IDENT (FP)
a. ☞e				*
b. a	!*	*	*	*

Candidate (b) violates all the low ranked constraints. It fatally violates CM-(MEAN) because it carries the meaning of the [Low] phoneme. Candidate (a) wins. It violates IDENT-(FP) because, overtly, it might be seen to represent other phones such as /e/ and /i/. However, it is the optimal candidate because it does not violate the higher ranked markedness constraint, CM-(MEAN).

In the representation of the long vowel /**ɛ**:/ in the word initial position, there are two candidates, <e> and <ee>. The presence of two candidates means that PI is infringed on and so MAX-(1P, 1G) constraints will be considered. Since <e> represents another phoneme, P2 is also contravened and IDENT-(1G, 1P) constraints must be considered. P3 is violated and so we consider IDENT-(FT) constraints. Length for the phoneme must not be represented in this case and so we consider *VIS-(V-L) constraints.

The ranking of the constraints is given below.

*VIS- (V-L) >> MAX (1P, 1G), IDENT (1G,1P), IDENT (FP)

The analysis is presented in the tableau below.

Tableau 20: Representation of /**ɛ:**/ in the word initial position

/**ɛ:**/ <e>	*VIS-(V-L)	MAX(1P, 1G)	IDENT(1G,1P)	IDENT(FP)
a. ☞e			*	*
b. ee	!*	*		

Candidate (b) is the loser. Though it does not violate IDENT-(FP) since the phone produced by speakers in this word position is a long one and length is ordinarily represented with double graphemes, it is not the graphemic representation of /ɛ:/ and so violates MAX-(1P, 1G). Its violation of *VIS-(V-L), which mitigates against representation of length in the word initial position, proves fatal and eliminates it. Candidate (a) is the winner. It violates IDENT-(FP) because it does not represent the phone for this particular phoneme. It also represents another phoneme, /e/ and so violates IDENT-(1G, 1P). These constraints are both ranked low. The candidate wins because it does not violate the high ranked markedness constraint, *VIS-(V-L).

In the medial and final word positions, there are three candidates, <ee>, <e> and <ea>. Their presence violates P1 and P2 and, therefore, MAX-(1P, 1G) constraints

must be considered. We also consider IDENT-(FP) constraints because all the candidates cannot be representatives of the phone spoken in this position. This is a contravention of P3. Additionally, length must be represented in these positions and therefore, we consider the VIS-(V-L) constraints. The different graphemes also mean that the meaning that is conveyed may not be the intended one, which is a violation of P4. Consequently, CM-(MEAN) constraints are considered.

The constraints used in the analysis are ranked thus:

VIS- (V-L) >> CM-(MEAN) >> MAX- (1P, 1G), IDENT-(1G,1P), IDENT-(FP)

The analysis is presented below.

Tableau 21: Representation of /ɛ:/ in the word medial and final positions

/**ɛ:**/ <ee>	VIS- (V-L)	CM- (MEAN)	MAX- (1P, 1G)	IDENT- (1G,1P)	IDENT- (FP)
a. ☞ee					
b. e	!*		*	*	*
c. ea		!*	*	*	*

Candidate (c) violates all the lower ranked constraints. It loses by violating CM-(MEAN) because the meaning it conveys is different from the meaning conveyed

by /ɛ:/. Candidate (b) violates IDENT-(1G, 1P) since it represents another phoneme. It also violates MAX-(1P, 1G) constraints because the phoneme is not represented by the grapheme. However, these are low ranked constraints. The candidate is eliminated because of violating VIS-(V-L) since it does not represent vowel length which must be shown in the medial and final word positions. Candidate (a) is the most harmonic. It does not violate any of the constraints.

The concatenation of this phoneme with other vocalic phonemes produces no discrepancies in the orthography. All the principles are observed and no constraint is violated.

4.5.8 The [+Low] Vowel /a/

This vowel was physically realized as [a] and had only one graphemic representation, <a> when it occurred as a short vowel. This is one case where all the principles under the phonological adequacy criterion were observed.

The grapheme for this vowel is easy to read and write both by native or non-native as well as by fluent or beginning speakers. The respondents are comfortable with its use and accept it in the graphemic inventory.

Discrepancies are noted with the representation of the long vowel /a:/. In the word initial position, the grapheme should be <a> but both <a> and <aa> are used. Consequently, P1 on accuracy and P2 on consistency are flouted and so MAX-(1P, 1G) and IDENT-(1G, 1P) constraints must be considered. Presenting <aa>

which represents vocalic length observes P3 (and therefore does not violate IDENT-(FP) but violates *VIS-(V-L) constraints. The constraints are ranked as shown below.

*VIS- (V-L) >> MAX- (1P, 1G), IDENT-(1G,1P), IDENT-(FP)

The analysis of the candidates is presented in the following tableau.

Tableau 22: Representation of /a:/ in word initial position

/a:/ <a>	*VIS- (V-L)	MAX- (1P, 1G)	IDENT- (1G,1P)	IDENT- (FP)
a. ☞a			*	*
b. aa	!*	*		

Candidate (b) represents vocalic length and so does not violate IDENT-(FP). It violates MAX-(1P, 1G) because it does not represent the phoneme in this word position. It fatally violates *VIS-(V-L) by representing vocalic length. This eliminates it. Candidate (a) is the winner. It does not represent the phone for the phoneme and so violates IDENT-(FP). It also represents another phoneme, /a/ and so violates IDENT-(1G, 1P). However, these are low ranked constraints. Since it does not violate *VIS(V-L), it emerges the winner.

In the word medial position, representation is at times with <aa> and other times with <a>. The principles, as well as the faithfulness constraints violated are the same as in the case of the phoneme in the word initial position. In the word medial and final positions, vocalic length must overtly be shown and so VIS-(V-L) constraints are considered. The constraints will be ranked as follows:

VIS- (V-L) >> MAX(1P, 1G), IDENT (1G,1P), IDENT (FP)

The analysis is shown in Tableau 23.

Tableau 23. Representation of /a:/ in word medial and final positions.

/a:/ <aa>	VIS- (V-L)	MAX- (1P, 1G)	IDENT (1G,1P)	IDENT (FP)
a. ☞aa				
b. a	!*	*	*	*

Candidate (b) is the loser. It violates all the constraints under consideration and so gets eliminated. Candidate (a) does not violate any constraint and so it wins.

The vowel /a:/ in concatenation with other vowels does not present any discrepancies.

4.5.9 The Back High [+ATR] Vowel /u/

The vowel /u/ was phonetically realized as [u]. There were two graphemic representations, <u> and <ũ>, and so, two candidates for evaluation. The grapheme <ũ> is the symbol for the mid back [+ATR] vowel /o/. The use of two graphemes reduces accuracy and thereby flouts P1. This is a violation of MAX-(1P, 1G). The two graphemes cause inconsistency by reducing predictability. This flouts P2, thereby violating MAX-(1P, 1G) and IDENT-(1G, 1P). P3, the Write as You Speak principle is flouted and so IDENT-(FP) is violated. The use of grapheme <ũ> for <u> could be explained as the representation of the gliding movement occasioned by a slight retraction of the tongue root and reduced auditory discrimination. Nevertheless, the sound produced is still [u] and not [o]. The consequence of using two graphemes for one phoneme is that the intended meaning is not clear. Ambiguity is created so that a word like ***u****ra*, 'to rain; to bleed' and ***ũ****ra*, 'get lost; run away' are deemed to have the same meanings. This flouts P4 and violates CM-(MEAN). On the other hand, the two symbols are in the Gĩkũyũ inventory and so P5 is not flouted.

For this phoneme, the ranking of the constraints is as follows.

> CM-(MEAN) >> MAX-(1P, 1G), IDENT-(1G, 1P), IDENT-(1G, 1P), IDENT-(FP)

The evaluation of the two candidates is shown in the tableau below.

Tableau 24: Representation of /u/

/u/ <u>	CM-(MEAN)	MAX-(1P, 1G)	IDENT-(1G, 1P)	IDENT-(1G,1P)	IDENT-(FP)
a. ☞u					
b. ũ	!*	*	*	*	*

Candidate (b) loses. It violates all the faithfulness constraints and fatally violates CM-(MEAN) because it conveys the unintended meaning. On the other hand, candidate (a) does not violate any of the constraints and so wins.

The representation of the long vowel /u:/ in the word initial position should be <u> but the respondents presented five graphemes which are <u>, <uu>, <ũ>, <i> and <ũu>. This shows inaccuracy and flouts P1 and is a violation of MAX-(1P, 1G) constraints. It is also inconsistent and unpredictable and so flouts P2, thus violating MAX-(1G, 1P). At the same time, grapheme <i> is a representation of a different phoneme and there is the insertion of a new element, [+ATR]. This violates DEP-(P, G). The phonemes do not represent the phones articulated for the phoneme. This contravenes P3 and violates IDENT-(FP). Looking at the graphemes, it is apparent one cannot say with certainity what phoneme is being

represented. This contravenes P4, the Meaning Conveyance principle and violates CM-(MEAN). All the graphemes used are found in the Gĩkũyũ inventory and so P5 has been observed. Some candidates also have a diacritic which makes them complex, something that violates *DIA.

The candidates can be ranked as given below.

> *VIS-(V-L), CM-(MEAN), *COMP >>MAX-(1P, 1G), DEP-(P, G), IDENT-(FP)

The analysis is presented in the following tableau.

Tableau 25: Representation of /u:/ in word initial position.

/u:/ <u>	*VIS-(V-L)	CM-(MEAN)	*COMP	MAX-(1P,1G)	DEP-(P, G)	IDENT-(FP)
a. ☞u						
b. uu	!*			*		
c. ũ		!*	*	*	*	*
d. i		!*		*	*	*
e. ũu		!*	*	*	*	*

Candidate (e) violates all the faithfulness constraints. The combination of two symbols renders it complex and so violates *COMP. It conveys an unintended meaning which makes it fatally violate CM-(MEAN). Candidate (d) fares much like Candidate (e) though it does not violate *COMP. It is eliminated when it fatality violates CM-(MEAN). Candidate (c) violates all the low ranking faithfulness constraints as well as *COMP. Its violation of CM-(MEAN) is, however, fatal. Candidate (b) violates MAX-(1P, 1G) because it is not the grapheme for the phoneme in the word initial position. However, this is a low ranked constraint in the language. It is knocked out of contention when it violates the *VIS-(V-L) constraint which mitigates against representing vowel length in the word initial position. Candidate (a) is the most harmonic and, therefore, the winner. It does not violate any of the constraints.

In the medial and final positions, four candidates were generated. These are <uu>, <u>, <ũ> and <ũu>. The undominated constraint, VIS-(V-L), which requires that vocalic length be shown, is at play. The other constraints remain the same as for the phoneme in the word initial position. Their ranking is shown below and the analysis given in Tableau 26.

VIS-(V-L), CM-(MEAN), *COMP >> MAX-(1P-1G), DEP-(P, G), IDENT-(FP)

Tableau 26: Representation of /u/ in word medial and final positions.

/u:/ <u>	VIS-(V-L)	CM-(MEAN)	*COMP	MAX-(1P,1G)	DEP-(P,G)	IDENT-(FP)
a. ☞uu						
b. u	!*			*		
c. ũ		!*	*	*	*	*
d. ũu		!*	*	*	*	*

Candidate (d) and (c) are the most disharmonic. They violate all the faithfulness constraints in addition to having diacritics, thereby becoming complex. They also convey unintended meanings and this eliminates them. Candidate (b) deletes a segment, [+long] thus violating MAX-(1P, 1G). It is eliminated because it does not represent length and so is a violation of VIS-(V-L), the highest ranked constraint. Candidate (a) is the optimal candidate. It wins by vitue of not violating any of the constraints under consideration.

The concatenation of this phoneme with others does not produce any discrepancies. All the principles and constraints are observed.

4.5.10 The Mid [+ATR] Vowel /o/

The vowel /o/ was phonetically realized as [o]. Three graphemic representations, <ũ>, <u> and <o> were proffered in all the word positions. The three flout P1

since only one should be in use. MAX-(1P, 1G) constraints are therefore, invoked. They also contravene P2 since there is no predictability as to which phoneme is being represented. The MAX-(1G, 1P) constraints therefore, become relevant. Speakers do not produce the vowels represented by all the graphemes and so P3 is flouted. Consequently, IDENT-(FP) constraints have to be considered. The intended meaning is also not clear when the three graphemes are used. As an example, the words *r**ũ**ga* 'jump', *r**u**ga* 'cook' and *r**o**ga* 'to poison' are seen to be synonymous. This flouts P4 and necessitates that we look at the CM-(MEAN) constraints. Nevertheless, the three are all in the Gĩkũyũ inventory and so P5 is not flouted. Consequently, we do not consider the INV-(G) constraints. The grapheme <ũ> is not used in both Kiswahili and English. To some Gĩkũyũ speakers, it is an unfamiliar symbol. This could explain why it is quite often not written and <u> is used instead. This contravenes P6. This calls for the consideration of FAM constraints.

The grapheme <ũ> is also not easy to use because of the diacritic. For this reason, writers use <u> in its place, whether in representing the short vowel, or doubling it to represent the long vowel, or in juxtaposing it with other vowels. This is because it is less cumbersome. Consequently, P7 is contravened and *DIA constraints are invoked. The use of <u> for <ũ> indicates that some of the users are uncomfortable with the grapheme and do not accept it. This is a flouting of both P8 and P9.

The constraints involved can be ranked as shown below. The evaluation of the candidates is presented in Tableau 27.

CM-(MEAN), *DIA >> FAM, MAX-(1P, 1G), MAX-(1G,1P), IDENT-(FP).

Tableau 27: Representation of /o/.

/o/ <ũ>	CM-(MEAN)	*DIA	FAM	MAX-(1P, 1G)	IDENT-(1G,1P)	IDENT-(FP)
a. ☞ũ		*	*			
b. u	!*			*	*	*
c. o	!*			*	*	*

Candidate (c) and (b) violate IDENT-(FT) because they represent phones that are not articulated for the phoneme. Candidate (c) deletes the property [+ATR] from the phoneme and inserts [-ATR] which violates the faithfulness constraints. Candidate (b) deletes the property [MID] and inserts [HIGH] which also violates the faithfulness constraints. Both candidates also violate CM-(MEAN) because they convey different meanings from what phoneme /o/ conveys. This is a fatal violation for the two. Candidate (a) is the winner. Though it violates FAM because it is a familiar symbol used in an unfamiliar way and also violates *DIA

because of the diacritic, it nevertheless, wins because it conveys the correct meaning carried by phoneme /o/.

In the representation of /o:/ in the word initial position, three candidates are generated, that is <ũ>, <ũũ> and <u>. The appearance of three graphemes for one phoneme flouts P1 and P2 and violates MAX-(1P, 1G), MAX-(1G, 1P), IDENT-(1G, 1P). Constraint DEP-(P, G) is also violated since one candidate represents a different phoneme. In this word position, vocalic length should not be shown so *VIS-(V-L) constraints are relevant. Graphemes should be easy to write and so *DIA constraints are relevant. They should also convey the intended meaning, hence CM-(MEAN) constraints are invoked. The ranking of the constraints is as shown below. The evaluation of the candidates is presented in Tableau 28.

*VIS-(V-L), CM-(MEAN), *DIA >> MAX-(1P, 1G), MAX-(1G, 1P),

DEP-(P, G), IDENT-(P, G)

Tableau 28: Representation of /o:/ in word initial position.

/o:/ <ũ>	*VIS- (V-L)	CM- (MEAN)	*DIA	MAX- (1P, 1G)	IDENT- (1G,1P)	DEP- (P, G)	IDENT- (FP)
a. ☞ũ			*				*
b. ũũ	!*		*	*	*		
c. u		!*		*	*	*	

Candidate (c) violates all the faithfulness constraints. Additionally, it fatally violates CM-(MEAN) because it conveys an unintended meaning. Candidate (b) violates both MAX-(1P, 1G) and IDENT-(1G, 1P) constraints but these are low ranked. However, its representation of vocalic length violates *VIS-(L-V) which eliminates it. Candidate (a) is the winner. It violates the low ranked IDENT-(FP) because it is not the physical representation of phoneme /o:/. It also violates the markedness constraint *DIA as it is not easy to write. All the same, it is the most optimal candidate because it fulfils the highest ranked constraint that calls for non-representation of vocalic length in the word initial position.

In the medial and final word positions, four candidates were generated. These were <ũũ>, <ũ>, <uu> and <u>. Multiple candidates flout P1 and P2 and violate MAX-(1P, 1G), IDENT-(1G, 1P) as well as DEP-(P, G). Some candidates also flout

P3 and violate IDENT-(FP), VIS-(V-L) and *DIA. The constraints are ranked as shown below:

VIS-(V-L), CM-(MEAN), *DIA >> MAX-(1P, 1G), IDENT-(1G,1P), DEP-(P, G), IDENT-(FP)

The evaluation of the candidates is presented in Tableau 29.

Tableau 29. Representation of /o:/ in Medial and final word positions.

/o:/ <ũ>	VIS- (V-L)	CM- (MEAN)	*DIA	MAX- (1P, 1G)	IDENT- (1G,1P)	DEP- (P, G)	IDENT- (FP)
a.☞ũũ			*				
b. ũ	!*		*	*	*	*	*
c. uu		!*		*	*	*	*
d. u		!*		*	*	*	*

Candidates (a), (b) and (c) violate all the faithfulness constraints. Both candidates (d) and (c) violate CM-(MEAN) because they convey the wrong meaning and are eliminated. Candidate (b) has a diacritic and so violates *DIA but her violation of VIS-(V-L) is fatal. This is so because it does not represent vocalic length.

Candidate (a) is the optimal one. It violates lower ranked markedness constraint *DIA but conveys the intended meaning and also represents vocalic length.

Concatenation of the vowel /o/ with /u/ and /ɔ/ creates no discrepancies and so contravenes no principle. When concatenated with /i/ to give <ũi>, additional graphemes <oi> and <i> are produced. Concatenation with /e/ to be represented with <ũĩ> also generates <ui> and <uĩ>. Concatenation with /a/ to be represented with <ũa> generates <ua>. Some respondents kept interchanging the graphemes <u> and <ũ> respectively. This switching and the additional graphemes flout the One phoneme – One grapheme Principle, disrupt the grapheme-phoneme correspondence and interfere with the conveyance of meaning. Thus, P1, P2, and P3 are flouted. P3 is also flouted since the speakers do not utter phones represented by these graphemes when they read their translations. Concatenation does not introduce new graphemes and so, P5 is not contravened. Apart from INV-(G), all the other constraints are violated in varying degrees.

4.5.11 The Back [-ATR] Vowel /ɔ/

This vowel is phonetically realized as [ɔ]. There were two graphemes used to represent it, <o> and <u>. P1 is flouted because only one grapheme ought to be used. This invokes the MAX-(1P, 1G) constraints. P2 is also flouted because <u> represents another phoneme, the back high [+ATR] /u/. Consequently, MAX-(1G, 1P) and DEP-(P, G) constraints are involved. P3 is flouted because there is only one physical realization to the phoneme and so IDENT-(FP) constraints come into

play. The intended meaning is lost and so P4 is flouted. This is based on CM-(MEAN) constraints. Since both graphemes are in the Gĩkũyũ inventory, P5 is not flouted.

The grapheme <o> for the back [-ATR] /ɔ/ presents no difficulties in regard to simplicity. Though it is used in English for /o/, /ɔ/ /and /u/ and only for /ɔ/ in Kiswahili, Gĩkũyũ speakers have no difficulties using it. The constraints under consideration for the phoneme can be ranked as follows.

CM-(MEAN) >> MAX-(1P, 1G), MAX-(1G, 1P), IDENT-(FP)

The analysis of the candidates is presented below.

Tableau 30: Representation of /ɔ/.

/ɔ/ <o>	CM-MEAN	MAX-(1P, 1G)	MAX-(1G,1P)	DEP-(P, G)	IDENT-(FP)
a. ☞o					
b. u	!*	*	*	*	*

Candidate (b) violates IDENT-(FT) because it does not represent the phone for the phoneme. It also violates DEP-(P, G) by inserting the feature [+ATR] and violates MAX-(1P, 1G) by deleting the feature [-ATR]. Fatally, it violates CM-(MEAN)

because it conveys an unintended meaning. Candidate (a) is the most harmonic since it violates none of the constraints.

Some difficulty arises in the representation of the long vowel /ɔ:/ in the word initial position. Some speakers double the representative grapheme, <o> as is the case in English in the word '**oo**ze'. Two candidates, <o> and <oo> are generated. This is a flouting of P1 and P2 and a violation of the MAX-(1P, 1G) and MAX-(1G, 1P) constraints. P3 is flouted in that, while some respondents represent the long vowel with the correct grapheme, <o>, others perceive length and so double the grapheme even though this is not the convention. This violates IDENT-(FP) as well as *VIS-(V-L). The intended meaning is conveyed and so P4 is observed. Since all the graphemes are in the inventory of Gĩkũyũ, P5 is not flouted.

The constraints for the evaluation of this phoneme can be ranked as follows.

*VIS-(V-L) >> MAX-(1P, 1G), IDENT-(1G, 1P), IDENT-(FP)

The evaluation of the two candidates is shown in Tableau 31.

Tableau 31: Representation of /ɔ/ in word initial position.

/ɔ:/ <o>	*VIS- (V-L)	MAX- (1P, 1G)	IDENT- (1G,1P)	IDENT-(FP)
a. ☞o				*
b. oo	!*	*	*	

Candidate (b) does not violate IDENT-(FP) because it represents the phone that is actually produced in the word position. However, it is not the grapheme for /ɔ/ and so violates IDENT-(1G, 1P) and also impinges on consistency thereby violating MAX-(1P, 1G). Gĩkũyũ orthography requires that vocalic length in the initial word position should not be shown and when (b) shows it, it fatally violates this constraint. Candidate (a) is the winner by not violating any of the constraints under consideration.

In the medial and final positions, the long vowel representation is by the same two graphemes representing the phoneme in the initial position. Similarly, the same principles, P1 and P2 are flouted while P3, P4 and P5 are observed. The faithfulness constraints at play are the same as those for the phoneme in the word initial position. A different markedness constraint is VIS-(V-L) that mitigates for visible vowel length in all positions except word initial. The constraints are ranked as follows:

VIS- (V-L) >> MAX-(1P, 1G), IDENT-(1G,1P), IDENT-(FP)

The evaluation is shown below.

Tableau 32: Representation of /ɔ/ in word medial and final positions.

/ɔ:/ <oo>	VIS- (V-L)	MAX- (1P, 1G)	IDENT- (1G,1P)	IDENT- (FP)
a. ☞oo				
b. o	!*	*	*	*

Candidate (b) violates all the faithfulness constraints. It is not the phone produced for the phoneme and it impinges on consistency in that it represents another phoneme. Moreover, it is not the grapheme for the phoneme. Its violation of VIS-(V-L) is fatal. Candidate (a) does not violate any of the constraints in contention. Concatenation of this phoneme with other vowels observes the adequacy criteria and so no discrepancies are observed.

4.6 Chapter Summary

In this chapter, we have presented the Gĩkũyũ phonemes and their attendant graphemes, both in the graphemic inventory and as generated by this study. With these, we have shown that discrepancies do exist in Gĩkũyũ orthography. We have discussed the extent of the discrepancies under three orthography criteria. We

have done this using the principles that underlie each criterion. From each of the principles, we have also identified the OT constraints. Using those constraints, we have analysed our data and accounted for the Gĩkũyũ graphemic structure using the Correspondence Theory.

One finding made in the chapter is that the advanced tongue root ([+/- ATR]), feature is best used for the description of the vowels in Gĩkũyũ. Another significant finding is that a good number of concatenated vocalic graphemes in Gĩkũyũ had not been identified and described prior to this study.

A major observation is that each of the orthography principles has been flouted by at least one grapheme. We have established that the extent to which the principles have been observed or flouted depends on the grapheme that is under analysis. We have also shown that a few graphemes are regular and do not flout any of the principles. We have accounted for these phenomena using Correspondence Theory constraints. According to Sasaki (2003), an efficient writing system should include a regular orthography, conventional and visually distinguished letters and there should be consistency between graphemes and phonemes. This is contrary to our findings. Our conclusion is that the regularity and irregularity observed in this chapter cast the Gĩkũyũ orthography as an inefficient writing system.

We have presented phonological explanations as to why the orthography principles are contravened by the graphemes. We further discuss these phonological phenomena in the following chapter and also examine the impact of the aforementioned discrepancies on the Gĩkũyũ speakers/ writers.

CHAPTER FIVE

UNDERREPRESENTATION AND OVERREPRESENTATION IN GĨKŨYŨ

5.0 Introduction

In the discussion about underrepresentation and overrepresentation, we need to first clarify and understand the concept of Orthographic Depth. We discuss this concept here so as to understand how underrepresentation and overrepresentation come about and their eventual impact on the speakers/ writers of Gĩkũyũ. We then discuss underrepresentation and overrepresentation for each of the Gĩkũyũ phonemes. We conclude the chapter with a summary.

5.1 Orthographic Depth

There are varied, though related, meanings to the concept of the depth of an orthography. According to Seifart (2006), the depth of an orthography refers to the level of linguistic structure at which the features represented in the orthography are located. It is about the directness and simplicity with which a writing system represents the phonology of a language (Katz & Frost, 1992). Orthographic depth is, therefore, the extent to which a grapheme is related to a phoneme. Two dimensions to the concept are thus identified: a shallow and a deep orthography.

A shallow orthography exhibits correspondence between a graphemic symbol and the physical realisation of the forms as they are pronounced (Seifart, 2006). This

means that a grapheme corresponds to a phone which is the physical realization of a phoneme. To take an example in Gĩkũyũ, if the physical realization of the alveolar nasal is [n], then the symbol that corresponds to that phone is used, which in this case is <n>. Additionally, each grapheme represents one phoneme, and each phoneme is represented by only one grapheme (Frost, 1989). Thus, from our example above of the alveolar nasal, <n> represents only phoneme /n/ and /n/ is represented by <n> only. This simply means that there is a one-to-one correspondence between a grapheme and a phoneme, a strict phonemic representation. A prime feature of this correspondence is that it creates ease in reading. In fact, as Hasselbring (2006) notes, a shallow orthography is primarily concerned with reading rather than writing.

On the other hand, a deep orthography exhibits correspondence between the grapheme and the underlying forms. In this case, a grapheme typically represents a distinct phoneme or phonemes but the correspondence is less direct. In effect, it means that the representation is less specific in respect to the phone(s) that it may hitherto represent. To give an example in English, grapheme <e> represents /i/, /e/ and /**ɛ**/. As observed, there are variations from a strictly phonemic representation. Karan (2006) observes that a deep orthography may result from underrepresentation or overrepresentation of phonemes, as well as from inconsistencies between grapheme-phoneme correspondences. Katz & Frost (1992) note that experienced readers do not process individual phonemes but, rather, longer strings and so are adept at reading deep orthographies. They

compensate for the orthography through comprehension of syntax, semantics and the context (Hasselbring, 2006).

In considering the two types of orthography, it is important to clarify that an outsider perceives the surface forms while a native speaker may even perceive phonological processes that he would want represented graphemically (Cahill and Karan, 2006). Irregular correspondence which leads to a deep orthography leads to difficulties in writing, reading, and decoding and recognising meaning. This research establishes the location of Gĩkũyũ in this deep or shallow dichotomy.

5.2 Underrepresentation and overrepresentation

Underrepresentation and overrepresentation are terms for variations from a purely phonemic representation (Hasselbring, 2006). The two are also referred to as underdifferentiation and overdifferentiation, respectively.

Cahill & Karan (2008) define underrepresentation as an instance where there are fewer graphemes than phonemes. In such a case, if a given language has X number of phonemes, the graphemes are X less 1 (or more). Karan (2006) adds that underrepresentation occurs when one grapheme represents more than one phoneme. For instance, from the data we collected in this study, <b> represents both /^{m}b/ and /β/, and this constitutes underrepresentation. Grapheme <b> will represent the two phonemes because the graphemes do not match the phonemes in number. In our research on the Gĩkũyũ graphemic structure, there is no scarcity of graphemes for specific phonemes yet underrepresentation occurs.

The converse of underrepresentation is overrepresentation. This occurs when there are more graphemes than there are phonemes. To illustrate this, if a given language has Y number of phonemes, the graphemes are Y plus 1(or more). The outcome of such a situation is the use of different graphemes for the same phoneme. A case in point is <c> and <s> representing /s/ in the words, *pea**c**e* and *ca**s**e* in English. In Gĩkũyũ, when some respondents present <b> and <mb> to represent /^{m}b/ in the words ***b**athi* (instead of *mbathi "bus"*) and *ng'o**mb**e* 'cow', this is overrepresentation.

Overrepresentation may also arise because of assigning different graphemes to allophones of a single phoneme. This is the case in Gĩkũyũ where respondents give <c>, <s> and <ch>to represent /ʃ/ in the word ***c**itũ*, ***s**itũ* and ***ch**itũ* (ours). One reason for this is dialectal variation where some dialects contrast sounds yet others do not. The consequence of overrepresentation is that the orthography becomes inconsistent though it may remain predictable, or it becomes both inconsistent and unpredictable (Sasaki, 2003).

Both underrepresentation and overrepresentation make the orthography deep. We discuss the phenomena of underrepresentation and overrepresentation with the view of establishing their genesis and the effect they have on the speakers/ writers of Gĩkũyũ. To do so, we shall examine each of the phonemes and their graphemic representations.

5.2.1 The Plosives and the Affricate

The voiced Gĩkũyũ plosives, the bilabial /^{m}b/, the alveolar /^{n}d/, and the velar /^{ŋ}g/ as well as the palatal affricate /^{ɲ}j/ are all preceded by the nasal segment /N/ that was discussed in § 4.3.1 and § 4.3.3. Armstrong (1967) observes that this nasal segment appears in a weak form in the word initial position and that in some cases, it is wholly lacking. The data we collected points to this fact where the segment in that position was almost lacking and some of the respondents failed to represent it graphemically thus causing overrepresentation. To some extent, we can attribute this non-representation to lack of perceptual distinction of the segment.

Additionally, in the Gĩkũyũ literature, there are many words where the nasal segment is unrepresented, for instance, in the Gĩkũyũ Bible. This is especially so with names of people and places, such as Daudi (David), Bethilehemu (Bethlehem), Goliathũ (Goliath) and Jesu (Jesus). Given that these are now familiar names in Gĩkũyũ, and that the speakers know them, we cannot ignore the influence their spelling might have on the orthography of our respondents. We therefore, add this as a reason for the non-representation of the nasal segment.

We also noted that English, which Gĩkũyũ highly interacts with, lacks the prenasalised phonemes and this could be a reason for the non-representation of the nasal.

The overall effect of having the nasal unrepresented is that the plosives and the affricate are rendered overrepresented where two symbols are used to represent one phoneme. In some cases, readers find some words strange and view them as not belonging to their orthography. Examples of such words are **b**egũ (instead of **mb**egũ ‘seed’), **d**ogo (for **nd**ogo, ‘smoke’), **g**arĩ (in place of **ng**arĩ, ‘leopard’) and **j**ũgũ (for **nj**ũgũ, ‘cow-peas’). In citation form, such words are foreign and do not have any meaning to the readers of Gĩkũyũ. In other cases, beginning readers, oblivious of any nasal segment that should precede the plosives and affricate read the words as they appear since some of the graphemes used are present in the phonetic inventory of Gĩkũyũ though they represent totally different phonemes. As examples, the word **mb**ata, ‘duck’ would be written as **b**ata and be read as [**β**ata] meaning ‘need’ while **ng**ũkũ ‘chicken’ would be written as **g**ũkũ [**ɣ**oko] which means ‘this place’. This interferes with the intended meaning.

On the other hand, competent native readers realise that the graphemes used represent the underlying sounds for the nasalised phonemes. Intuitively, they recognise the phonological processes that the phonemes have gone through and identify the graphemes representing them. They not only recognise the words that are deemed ‘strange’ by beginner readers, but also read them correctly. In a phonological phrase, context also aids them in interpreting the intended meaning of words (Perry, Ziegler and Colheart, 2002; Venezky, 2004). For example, if the context is on transport, the grapheme <b> in **b**athi ‘iron’ will be accurately

deciphered as <mb> in **mb**athi ‘bus’ and the intended meaning will be understood.

5.2.2 The Nasals

As noted from our data, the bilabial nasal /m/ and the alveolar nasal /n/ are neither underrepresented nor overrepresented. Consequently, the two do not present any writing or reading complications to Gĩkũyũ speakers.

The palatal nasal /ɲ/ has one surface realisation [ɲ] but two graphemic representations, <ny> and <nj>. The grapheme <nj> is presented by speakers from Mathĩra and Ndia. This grapheme causes underrepresentation of the phoneme because it is also used to represent the prenasalised affricate /^{n}ɟ/. Aside from /ɲ/ and /^{n}ɟ/ sharing the same place of articulation, it is worth noting that /ɲ/ is the nasal that the affricate /ɟ/ takes after the place assimilation of /N/. The phonological closeness of the two explains why some speakers cannot perceptually discriminate the two and so present them as one graphemically. The closeness might also explain why the differentiation of the two phonemes is so important that even exact minimal pairs are hard to find in Gĩkũyũ. The meaning conveyed by these phonemes is distinct and readers get confused where the graphemes are used interchangeably.

The velar nasal /ŋ/ also has a single surface realisation [ŋ] and two graphemic representations <ng’> and <ng>.The second grapheme causes underrepresentation because it is used to represent /^{ŋ}g/. We note that this grapheme is used by

respondents from all dialects except Mathĩra who produce it correctly as <ng'>. The significance here is that Mathĩra speakers might be more careful in their writing with the intention of keeping the phoneme distinctly represented. Overall, the diacritic is usually abandoned in spontaneous writing because it hinders fluency. In reading, the diacritic requires more effort to be identified accurately due to the fact that it is not very salient (Baroni, 2016). Readers from the other dialects leave the distinction to context. This portends difficulties in distinguishing meaning.

5.2.3 The Fricatives

The Gĩkũyũ fricatives are identified as the voiced bilabial /β/, the voiced inter-dental /ð/, the voiceless palatal /ʃ/, the voiced velar /ɣ/ and the voiceless glottal /h/ and their respective graphemic representations are <b>, <th>, <c>, <g> and <h>. The graphemes produced by our respondents create a lot of both underrepresentation and overrepresentation.

Starting with phoneme /β/, it is represented graphemically by <b> and is articulated variously by Gĩkũyũ speakers as [β], [ɸ], [v], [f] and [p]. Phones [β] and [ɸ] are both bilabial fricatives and only differ in that [β] is voiced while [ɸ] is voiceless. Phones [v] and [f] are labial dental fricatives while [p] is a bilabial plosive. That the first two are labial dentals and [p] is a plosive makes these three quite distinct from the bilabial fricatives, yet they are produced in place of /β/. Unfortunately, while it seems that there is consensus for <v>, <f> and <p> to

represent /v/, /f/ and /p/ respectively, the same does not seem to hold for either [β] or [ɸ] (Sampson, 2016; Wiese, 2004). In any case, each speaker has the intention of representing the phone that he perceives for that phoneme, so <b>, <mb>, <p>, <v> and <f> are all used. This explains the overrepresentation of this phoneme. Grapheme <mb> also brings about underrepresentation because it is used to represent /^{m}b/.

Grapheme <b> is invariably used in English and Kiswahili for phoneme /b/. The two languages are in constant interaction with Gĩkũyũ and the Gĩkũyũ speakers would want to create maximum differentiation between their /β/ and the /b/ of these two languages. For instance, given the word 'baba', which in English is pronounced as [ba:ba:] and in Kiswahili as [baba] ('father'), we assume that Gĩkũyũ speakers would want to distinguish it from [βaβa] (father). Consequently, they avoid using <b> and instead use <mb>, <p>, <v> and <f>, thereby causing overrepresentation of the phoneme.

Where underrepresentation is caused by <mb>, the readers may face the problem of determining the phoneme that is represented. This may necessitate getting clues from the context, something that would obviously interfere with the reading speed. Overrepresentation would require that the readers mentally recognise <b>, <mb>, <p>, <v> and <f> as all representing /β/ which is an ardours task in reading. This may explain why 37% of the respondents find reading in Gĩkũyũ

difficult. The overrepresentation may also explain why some respondents would want <f> included in the graphemic inventory as the representation of /β/.

The voiced inter-dental fricative /ð/, whose grapheme is <th>, is represented graphemically with <nd>, <d> and <ny> by some speakers from Ndia and Gĩcũgũ. Some of the respondents articulate the phoneme as the voiced alveolar [z] which they possibly perceive as the alveolar voiced plosive [d] and represent it as <d>. Other respondents articulate the phoneme just behind the dentals as [d̪], dentalised, and not really as an inter-dental. This is closest to the pre-nasalised /^{n}d/ and so they represent it with <nd>. The pre-nasalised plosive /^{n}d/ has gone through the weakening process in Gĩkũyũ to be realised as /ð/. With further weakening, it might be realised as /θ/. We submit that in Ndia and Gĩcũgũ, /^{n}d/, through weakening, is yet to be realised as /ð/, hence the representation with <nd>.

It might be that other respondents do not correlate /ð/ with <th> because of the ‘h’ element in the grapheme which, to them, represents a phoneme that they do not possess. They avoid the ‘h’ and are left with the ‘t’ segment. This segment corresponds to the alveolar where the voiceless /t/ is produced. Possibly, because they can perceive voice, they switch the /t/ with the voiced alveolar /d/ and then use <d> to represent the phoneme. This would cause overrepresentation and introduce a new grapheme.

The underrepresentation caused by <nd> creates confusion and makes reading difficult even for the speakers themselves. Some are unable to decipher what they have written and this interferes with their speed of reading. Other speakers read the graphemes <d> and <nd> as they are presented and rely on the phonological phrase to get the meaning. This, of course, interferes with accuracy in reading. Grapheme <th> demands extra effort to master so that it can be correlated with its phoneme. For this reason, some speakers would rather have it excluded from the graphemic inventory and indeed, do suggest so.

The voiceless palatal fricative /ʃ/ is overrepresented with <c>, <s> and <ch>. Speakers from the Southern, Ndia and Gĩcũgũ dialects articulate [s] and [ʃ] for the phoneme while [tʃ] is articulated by speakers from the Northern and Mathĩra dialects. The speakers from the Northern and Southern dialects were representing the sounds they articulate for the phoneme. In addition, they could have been representing the sound the same way it is done in both English (for instance in **s**ay and **ch**ip) and in Kiswahili (for example in **s**ema, 'say' and **ch**ai 'tea'). The grapheme <c> demands extra concentration especially to the speakers who are more comfortable using English than Gĩkũyũ since it represents other phonemes such as /k/ in **c**ar and /s/ in **c**ease. This could possibly explain why some speakers suggest that the grapheme <s> be used instead, so as to ease the burden of both writing and reading.

The velar fricative /ɣ/ is represented with both <g> and <k>. The use of <k> is an instance of underrepresentation since the same grapheme represents /k/. Phonemes /g/, /ᵑg/, /k/ and /ɣ/ are produced at the same place, the velar. While /g/ does not exist in Gĩkũyũ, the underrepresentation with <k> may be explained by the process of consonant weakening. In this process, the voiced phoneme /ᵑg/ is weakened to the voiceless /k/ which is further weakened to the fricative /ɣ/. The weakening process is thus;

/ᵑg/→ /k/→/ɣ/.

The arrow shows the direction of the process.

Since the representation of /k/ is the grapheme <k> that all our respondents presented, we submit that phoneme /ᵑg/ is still going through the weakening process and that is why some respondents from Gĩcũgũ and Ndia used <k> instead. This is borne out by the fact that these speakers also produce the phone [k] for /ɣ/.

Some readers may not find reading the graphemes <g> and <k> difficult because the phonemes they represent are closely allied, differing only in the aspects of either voice or manner of articulation. On the other hand, readers outside Gĩcũgũ and Ndia may get confused because different lexical items are assigned the same meaning. For instance, the minimal pair, ‘ri**k**a’ /ɾika/ (dive into/ settle) and ‘ri**g**a’ /ɾiɣa/ (be unknown) are now seen as a single lexical entity. Additionally, any

other word with either of the graphemes would need greater effort in deciphering the meanings.

The voiced glottal fricative /h/ is also represented with <b> and <mb> by some respondents from Ndia and Gĩcũgũ, causing underrepresentation of the phoneme. This is not just a case of fronting and consonant switching but one where grammatical tone is at work. When <b> represents /β/, the speakers use a falling tone but when it represents /h/, they employ a rising tone. Since [h] is less sonorous than [β] and the two phones merge into [β] in these dialects, differentiation is achieved through raising the sonority in the [β] that represents /h/. Thus, **'bùbù'** (a deaf and dumb person) and **'búbú'** (a bat) are differentiated. However, tone is not represented orthographically in Gĩkũyũ. The speakers in these dialects are able to differentiate between the <b> that represents /h/ and the one that represents /β/ but this situation does not hold with the speakers of the other dialects.

Given the context, readers from the Ndia and Gĩcũgũ dialects will disambiguate the words. Readers from the other dialects will not discriminate since <b> represents only one phoneme. The result is that the intended meaning is lost to them.

5.2.4 The Alveolar Tap /ɾ/

The tap /ɾ/ and the lateral /l/ are liquids and are acoustically similar (Fromkin, 2007). In Gĩkũyũ, the phoneme is produced with a single tap of the tongue against

the alveolar. Therefore, it is neither a roll nor a trill (Benson, 1964; Armstrong, 1967). The two phones, [ɾ] and [l] are allophones of /ɾ/ and occur in free variation (Benson, 1964). This phonemic non-distinction is carried across to graphemic representation and is a cause of overrepresentation.

The fact that these phonemes are allophones means that native Gĩkũyũ speakers do not find any difficulties in reading for they are able to easily infer the intended meaning. It is only with non-native speakers (who distinguish between the phonemes in their languages) that problems would arise where the meaning for /l/ would be looked for in vain.

5.2.4 The Short Vowels

Underrepresentation and overrepresentation are also noted with the vowels. We analyse them in the following sub-sections.

5.2.4.1 The Front [+ATR] Vowels /i/ and /e/

The front [+ATR] phonemes are graphemically represented with <i> and <î>, respectively. Some Gĩkũyũ speakers interchange the graphemes, causing underrepresentation. Both phonemes /i/ and /e/ are produced with an advanced tongue root and so have the characteristic‘breathy’, where a greater deal of air passes through a slightly open glottis. Phoneme /i/ is more ‘breathy’ than /e/ and requires more effort. Since an articulation which requires more effort is disfavoured (Boersma 1998, p 149), we argue that this is the reason their attendant graphemes cause confusion and why they are consequently exchanged.

In addition, the characteristic ‘breathy’ may not be contrastive enough for the two vowels in Gĩkũyũ. Consequently, some writers represent the two with any of the graphemes designated for the two phonemes.

The result of the interchange of graphemes as explained above is that Gĩkũyũ competent readers do not face great difficulties in reading them. Nevertheless, there are some hesitations while processing the graphemes to retrieve meaning.

As noted in our discussion elsewhere, the front [+ATR] vowels do not mix with the front [-ATR] /ɛ/ as observed by Peng (2000). The perceptual distinction between the two groups of vowels is big enough for one not to be taken for the other. This explains why no writer interchanges the graphemes <i> and <ĩ> with <e>, the grapheme for the front [-ATR] /ɛ/.

5.2.4.2 The Front [-ATR] Vowel /ɛ/

The front [-ATR] phoneme is underrepresented in that, apart from representing it with <e>, some writers also present <a> for it. Phoneme /ɛ/ is produced with the tongue root retracted while phoneme /a/, the vowel next to it, is produced with the tongue root in the neutral position. On the other hand, phoneme /e/, the other vowel neighbouring /ɛ/, is produced with the tongue root advanced. Due to minimization of extension, whereby an articulator prefers to stay as near as possible to the neutral position (Boersma, 1998), the speakers perceive the phoneme /ɛ/ as the neutral /a/ and represent it as <a>. The writers are therefore, representing this phonological reality when they use this grapheme. For readers,

confusion is created and more effort is required to decipher meaning unless context is taken into account. Such would be the case, for example, between the words *c**e**ma* (to stalk) and *c**a**ma* (flavour/ sweetness).

5.2.4.3 The Back [-ATR] Vowel /ɔ/

The back [-ATR] vowel /ɔ/ was represented by both <o> and <u>. The second grapheme was used by a speaker from the Southern dialect. This grapheme represents the back [+ATR] vowel /u/. Yet, as we had noted earlier, [+ATR] and [-ATR] vowels do not interchange. This situation may be explained by the influence of the English orthography on the speaker. In some English words, for instance 'fool' and 'tool', grapheme <o> is used to represent /u/. This might have caused perceptual confusion to the speakers. Instead of writing <o>, they write <u> as they try to maximally distinguish phoneme /ɔ/ from phoneme /u/. This creates underrepresentation which causes difficulties to readers as they try to retrieve phonological meaning from the grapheme. A word such as mb**o**gi 'goods wagon' written as mb**u**gi 'a bell-shaped metal object' presents a real challenge to the information flow.

5.2.4.4 The Back [+ATR] Vowels /o/ and /u/

The graphemes for the two back [+ATR] vowels /o/ and /u/ are <ũ> and <u>, respectively. In addition to the two, some respondents also use <o> to represent /o/. The fact that the two vowels share similarity in the position of the tongue root position may explain why some writers use the graphemes designed for the

phonemes interchangeably. It may also explain why to the native speakers, the use of one grapheme for the other does not present real reading problems. Hidden behind this explanation is the likelihood that the writers are gradually merging the two graphemes, especially given that <ũ> is more cumbersome to write because of the diacritic.

In addition to <ũ>, some respondents also use <o> to represent /o/. This representation of /o/ with <o> could be attributed to the influence of English where <o> is used, for instance, in the words ‘old’ and ‘coat’. It could also be attributed to confusion created by the similarity in visual perception of /o/ and <o>. The use of <o> instead of <ũ> in Gĩkũyũ creates problems to the readers because it interferes with the intended meaning. As an example, it is hard to decide what the writer intends between the words m**ũ**nd**ũ** (person) and m**o**nd**o** (bag) outside context.

5.2.5 The Long Vowels

Vowel length is indicated by the doubling of the vocalic grapheme in all word positions except the initial position. In all the positions, underrepresentation is caused by either the use of doubled graphemes or graphemes that represent other phonemes.

5.2.5.1 Long Vowels in Initial Word Position

In the initial word position, respondents have used both the single grapheme as well as double graphemes. This creates overrepresentation. Graphemes for other

phonemes are also used which causes underrepresentation. The front mid [+ATR] vowel /e:/, which should be represented with /ĩ/, is additionally represented with <i> and with <ii> that should otherwise represent /i:/. We present the argument of the 'breathy' quality as not being distinctive enough, for the interchange of the graphemes. We note that the front [+ATR] graphemes are not interchanged with the front [-ATR] grapheme. From this observation, we posit that the writers/ readers are aware of the phonology of their language and that they strive to represent it graphemically. Consequently, the underrepresentation of /i:/, /e:/, /ɛ:/ and /a:/ do not present difficulties to the Gĩkũyũ readers or speakers and meaning is easily retrieved.

With the back [-ATR] vowel /ɔ:/, the double grapheme <oo> is also used. Again, the writers/ readers are representing the phonology of the language as they know it. Graphemes for the long back [+ATR] vowels, that is <u> and <ũ>, are sometimes doubled and interchanged. For the two phonemes, we argue that the 'breathy' quality does not help in distinguishing them and that the readers are presenting their phonology. In all cases, there is no interference with reading; and meaning is derived from context. However, reading and retrieval of meaning is harder where <i>, the grapheme for the front high [+ATR] vowel, is used for the back high [+ATR] vowel, <u>.

5.2.5.2 Long Vowels in Medial and Final word positions

In the medial and final word positions, the use of doubled vocalic graphemes indicates vowel length. However, in all representations, more than one grapheme is presented by the respondents. This creates overrepresentation of the phonemes. Every phoneme is also represented with the grapheme that represents its short counterpart. This is underrepresentation. In other cases, underrepresentation is seen where a grapheme that represents a totally different vowel is used. This includes <ii> and <i> being used for /e:/, <ũ> for /u:/ and /u/ for /o:/. Underrepresentation is also seen where vowels in concatenation are used. This is the case with <ea> for /ɛ/, <ĩi> for /e:/ and /ũu/ for /u:/.

The causes for both underrepresentation and overrepresentation are the same as those advanced for long vowels in the initial word position. The consequences are also the same and include interference with fluency in reading and difficulties in retrieving the intended meaning.

5.2.6 Vowels in Concatenation

Underrepresentation and overrepresentation also occur with vowels in concatenation. Of note from the data is that the vowels that undergo the phenomena map neatly onto the short vowels noted in § 5.4.2.1. The table on the next page summarizes the occurrence.

Table 5. 1: Concatenated Vowels and their Graphemic Representation

Phoneme	Grapheme	Variant Graphemes			
/ie/	<iĩ>	<ii>			
/io/	<ĩũ>	<iu>	<ĩũ>	<io>	<ĩu>
/ia/	<ia>	<ĩa>			
/ei/	<ĩi>	<ii>			
/eo/	<ĩũ>	<ĩu>	<iu>		
/eɛ/	<ĩe>	<ie>			
/eɔ/	<ĩo>	<io>			
/ea/	<ĩa>	<ia>			
/oi/	<ũi>	<oi>	<i>		
/oe/	<ũĩ>	<ũi>	<uĩ>		
/oa/	<ũa>	<ua>			

As a way of explaining the phenomena of underrepresentation and overrepresentation, we advance the same reasons given in §5.4.2.1 – §5.4.2.5. The reasons include: similarity in tongue root position for phonemes /i/ and /e/ on one

hand, and /u/ and /o/ on the other (resulting in free variation; ease of articulation with the same pair of phonemes where one phoneme is more easily produced and transfer to writing fails; lack of perceptual distinctiveness of the 'breathy' quality between /i/ and /e/; and, fronting of /o/ to /i/. The phenomena hamper reading because hesitations and sub-vocalization are bound occur in reading. Retrieval of meaning from the graphemes is also rendered difficult. Overall, rapid reading becomes near impossible.

Native Gĩkũyũ speakers/ readers would have no problems with deducing the correct meaning because they are intuitively aware of the phonotactics of the language. The tone that is assigned to the vowel and the context do help in minimizing perceptual confusion. For instance, if the written word is '*mai'*, the /i/ may be assigned a rising tone to give ma**i** 'faeces' or be given a falling tone which can get to as low as /e/ with the word being vocalised as *maĩ* **'**water'. But as noted, this will depend on the context.

5.3 Chapter Summary

In this chapter, we have discussed underrepresentation and overrepresentation in Gĩkũyũ and how the two relate to orthographic depth. We have examined the factors that lead to the two and how they impact on Gĩkũyũ speakers and writers. We have not discussed the approximants because they do not exhibit the phenomena of underrepresentation and overrepresentation.

Our conclusion is that the irregular and unpredictable correspondence between the phoneme and the grapheme that result from both underrepresentation and overrepresentation, leads to a deeper orthography. Correlating graphemic information to phonological information is rendered more difficult given that accuracy and consistency are hampered. Moreover, efficiency and ease in writing and in reading, as well as in the retrieval of semantic information, are interfered with. This makes Gĩkũyũ difficult to read and write.

In this study, though the subjects were expected to be fluent in writing and reading Gĩkũyũ, they still displayed discrepancies in their writing. This can be attributed to their endeavour at mastering the writing system by reading of whole words as opposed to individual phonemes. This is a strategy of handling a deep orthography. This observation concurs with Seifart (2006, p. 282) who notes that from the perspective of psycholinguistics, 'the optional orthography for a beginning reader is not the same as for a fluent reader'. Advanced readers heavily rely on "sight vocabulary"; that is, written words are recognised as entire units and processed as such, without breaking them down into units of the sound structure.

CHAPTER SIX

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

6.0 Introduction

This study aimed at explaining and accounting for the discrepancies in the Gĩkũyũ graphemic structure using the Correspondence Theory, a sub theory of the Optimality Theory. Towards this endeavour, three questions were raised and their answers sought. The first question addressed the extent to which the criteria for designing an orthography are observed in the orthography of Gĩkũyũ. The second question examined how the Correspondence Theory of the Optimality Theory can account for the Gĩkũyũ graphemic structure. The last question examined the impact of underrepresentation and overrepresentation of the orthographic criteria on the writers and readers of Gĩkũyũ. This chapter presents a summary of the findings based on these three research questions. It also gives conclusions based on the findings. These are followed by recommendations for general linguistics and for the theory, as well as areas for further research.

6.1 Summary of Findings

To address our research questions, we interrogated the Gĩkũyũ phonemes and the graphemes that the study generated. The phonemes were eighteen consonants, seven short vowels, seven long vowels, and thirty-four combinations of vocalic phonemes which we referred to as vowels in concatenation.

Amongst our first findings is that the feature [+/- ATR], as opposed to horizontal tongue position or height, is the best for describing Gĩkũyũ vowels. This feature aptly describes the vowels of African languages. It is especially appropriate in describing the Gĩkũyũ vowels which have been described as lacking the crispness and precision of the English vowels (Armstrong, 1967, p. 4).

A major finding is with regard to the number of concatenated vowels in Gĩkũyũ. Only twenty - four such vowels have been subjected to study in previous works (Armstrong, 1967, p. 2). We were able to identify ten other concatenations and subject them to analysis. These are /eu/, /ɛɔ/, /ɛa/, /ou/, /ɔɛ/, /ɔa/, /aɔ/, /ao/, /au/ and /ae/. Their graphemic representations are <ĩu>, <eo>, <ea>, <ũu>, <oe>, <oa>, <ao>, <aũ>, <au> and <aĩ>, respectively. Out of these thirty-four, only seven are found in the word initial position. These are /ai/, /aɔ/, /ao/, /ɔi/, /ɔe/, /oi/ and /oe/. Vowels /iu/, /eo/, and /ɛi/ are not found in the word final position. The vowel /ɔa/ was present in the word final position in only one word (ng'oa, 'assemble') that is not in common use.

We made a major finding concerning OT constraints emanating from the first and second research questions, We were able to identify, hitherto, unidentified constraints that can be used in other graphonological studies. In addressing the two objectives, we segmented the criteria for designing an orthography into the constituent principles. From the principles, we extracted faithfulness and markedness OT constraints under which each principle operates (Wiese & Song,

2010). The faithfulness constraints include MAX- (1P, 1G), IDENT- (1G, 1P), DEP- (P, G), IDENT- (F, P), IDENT-(FT), FAM and ACCEPT. The markedness constraints are NC- (N, P), VISIBILITY (which we split into VIS-(V-L) and DIA-(F)), CM- (MEAN), *COMP, *DIA and INV-(G). We then subjected the data to analysis against each of the principles and the relevant constraints.

From the analysis, we found that six consonantal phonemes do not contravene the principles or violate any constraint in their graphemic representation. The six are the voiceless alveolar plosive /t/; the voiceless velar plosive /k/; the bilabial nasal /m/; the alveolar nasal /n/; the bilabial approximant /w/; and the palatal approximant /j/. We summarise the findings for the other phonemes in the following paragraphs.

The phonemes contravene P1, the One phoneme - One grapheme correspondence. The constraints underlying this principle are the faithfulness MAX: (1P, 1G) and the markedness NC: (N, P). By representing one phoneme with two or more graphemes, the outputs violate the MAX: (1P, 1G). This constraint is, however, low ranked in the language and does not serve to eliminate candidates from deciding the most harmonic of the candidates. Additionally, some of the outputs for pre-nasalised phonemes violate the markedness constraint because they delete the nasal segment in the phoneme and so do not represent it graphemically. In Gĩkũyũ, this is a fatal violation that eliminates those candidates.

The second principle, P2, the One grapheme - One phoneme principle, requires that one grapheme represents a single phoneme. This creates consistency and also mitigates against insertion. In the representation of consonants, grapheme <b> represents /β/, /^{m}b/, and /h/ while <mb> represents /mb/ and /h/. Grapheme <nd> represents /^{n}d/ and /ð/. Grapheme <k> is given to represent /k/ and /ɣ/ while <g> represents /ɣ/ and /^{ŋ}g/. Grapheme <ng> represents /^{ŋ}g/ and /ŋ/ while <ng'> represents /ŋ/and /^{ŋ}g/. Grapheme <ny> represents /ŋ/ and /ð/ while <nj> represents /^{n}ɟ/ and /ɲ/. In the representation of vowels, grapheme <a> represents /a/, /a:/ and /ɛ/ while <aa> represents /a:/ and /a/. Grapheme <i> represents /i/, /i:/ and /e/ while <ii> represents /i:/ and /i/. Grapheme <ĩ> represents /e/, /e:/, /i/ and /i:/ while <ĩĩ> represents /e:/ and /e/. Grapheme <e> represents /ɛ/, /ɛ:/ and /a/ while <ee> also represents /ɛ:/ and /ɛ/. Grapheme <o> represents /ɔ/, /ɔ:/ and /u/ while <oo> represents /ɔ:/ and /ɔ/. The grapheme for /o/ which is /ũ/ also represents /o:/, /u/ and /ɔ/ while the grapheme for /o:/, that is /ũũ/ also represents /o/.

The relevant constraints identified under P2 were the faithfulness MAX- (1G, 1P), IDENT- (1G, 1P), and DEP- (P, G). All the graphemic outputs identified in the paragraph above violate these constraints. Fortunately, the constraints are ranked low in Gĩkũyũ and their violation does not serve to eliminate the candidates from competing for optimality.

The third Principle, P3 is the Write as You Speak principle that requires representation of phones. To a great extent, this principle has been observed. This explains the proliferation of diverse graphemes for a single phoneme as speakers represent their idiosyncrasies through writing. We referred to these graphemes as allographs. We observed that different speakers produce different phones (allophones) for a phoneme and represent the phoneme with a grapheme that best approximates the particular phone. For example, a speaker who produces [l] instead of [r] may represent the sound with <l>, and the one who produces a plosive without the prenasal may represent it graphemically without representing that segment. We attribute this to idiosyncrasies, dialect and language evolution, especially in the case of the dropping of the prenasalised segment.

The faithfulness constraint we identified here was IDENT- (F, P). We noted that this constraint ranks lowest in Gĩkũyũ and so its violation had no impact on the outcome from the constraint evaluation. We identified two markedness constraints, VISIBILITY which we split into VIS-(V-L) and DIA-(F) where the former, whether in its positive or negative form, ranks higher than the latter. This makes it a decisive constraint. In the representation of the velar nasal, the candidate which did not visibly show that the input was a nasal (by use of the apostrophe) violated VIS-(V-L) and so, was eliminated. With the vowels, there should be no representation of vocalic length in the initial word position and so, *VIS-(V-L) is decisive. Graphemes that indicated length were thus eliminated. In the medial and final word positions, vowel length must be indicated. For this

reason, VIS-(V-L) is decisive. Graphemes that did not indicate length were eliminated.

P4 is the Convey Meaning Principle which requires that a grapheme conveys the meaning that a phoneme intends. Where there is more than one grapheme for one phoneme, or one grapheme represents more than phoneme, this principle is contravened. We observe that this was the case with all the outputs except <m>, <n>, <w> and <y>. From the principle, we derived the markedness constraints CM- (MEAN). The constraints were decisive in many cases where they were used to account for graphemes that were found to carry meanings that differed from those intended by the phoneme.

The next principle, P5 requires that only those graphemes that are found in the inventory of a language be used. We established that graphemes <d>, <j>, <f>, <v>, <p>, <l>, <s>, and <ch> are not in the inventory of Gĩkũyũ and, therefore, contravene this principle. We derived the INV-(G) constraints from this principle. These are the highest ranked constraints in the language and their violation is fatal. The eight outputs given above violated these constraints and were found to be disharmonic.

The next principle P6, is the Principle of Overall Least Effort. Use of graphemes that have diacritics, or are a cluster of symbols, or are used differently in interacting languages, require more effort and are likely to cause discrepancies.

Related to P6 is the Easiest for Users principle, P7. It calls for salience in visual appearance and ease in writing. Diacritics and digraphs lead to contravention of the principle. The constraints that we identified from P6 and P7 were the faithfulness IDENT-(FT) and the markedness *COMP and *DIA which are not decisive.

The next two principles are P8, the Familiarity Principle which calls for familiarity with the graphemes that are used, and P9, the Acceptance of Grapheme Principle. We established that some speakers are more familiar with <b>, <d>, <g> and <j> for the pre-nasals than with the designated graphemes. Some respondents reject <ng'> especially because of the apostrophe, thus contravening P9. From these principles, we identified two faithfulness constraints, FAM and ACCEPT.

We were able to establish that Gĩkũyũ orthography is a deep orthography on the basis of the third research question addressed by this research. The question interrogated the impact of underrepresentation and overrepresentation on the writers and readers of Gĩkũyũ. We observed that both phenomena exist in varying degrees. While they are not found with /t/, /m/, /j/, /w/ and /a/, they are found with all the other phonemes of Gĩkũyũ. As to the causes for the underrepresentation and overrepresentation, we advanced graphonological, orthographic, and other non - linguistic factors. The phonological factors we identified include: lack of perceptual distinction of segments; consonant switching; ongoing phonological

processes such as weakening of phonemes; grammatical tone that interferes with vocalic length; allophones, for example [s], [ʃ] and [tʃ]; acoustic similarity as is the case with /r/ and /l/; the quality ‘breathy’ of vocalic phonemes; and, minimalization of effort where, for example, an articulator (the tongue) prefers staying in a neutral position. Orthographic factors include the fact that there are no universally agreed on graphemes for some phonemes such as /ɸ/ and /β/, and that others are complex to write. Non - linguistic factors include borrowing of graphemes, familiarity with graphemes used in other languages, and lastly, the desire by speakers/ writers to create distinctiveness so that Gĩkũyũ orthography is seen as different from that of other languages.

Additionally, we have established that the underrepresentation and overrepresentation cause confusion leading to difficulties in processing graphemic information. The consequence is inability to extract the intended meaning. This is countered by, first, taking recourse in context so as to get meanings. Secondly, speakers use their phonotactic knowledge of Gĩkũyũ. Another consequence of the two phenomena is impeded fluency in reading and difficulties in writing.

6.2 Conclusions

To conclude, we have established that the Principles of Orthography are contravened to a large extent by the speakers/ writers of Gĩkũyũ, and to some extent by the designers of the orthography. Additionally, we have shown how the contraventions have come by and have suggested some remedies. We have

applied the tenets of the Correspondence Theory and shown that the theory is capable of handling graphonological phenomena. Lastly, we have identified the consequences of underrepresentation and overrepresentation on Gĩkũyũ speakers/ writers and discussed the counter measures that the speakers take.

6.3 Recommendations

The study has identified graphemes that native Gĩkũyũ speakers feel are representative of their phonemes. These include <s> for /s/ and <f> for /β/. We recommend that a study be done with the aim of revising the orthography of Gĩkũyũ so as to include these graphemes.

This study has observed that the feature [ATR], which has been used to describe vowels in other Bantu languages, is the best way of describing Gĩkũyũ vowels. Such a description can help in accounting for the vowel switching that has been observed in the study. Consequently, we recommend that a phonetic study be done to unequivocally establish the vowel space in Gĩkũyũ. Additionally, we recommend that strategies used by speakers of languages with more than five vowels but five vocalic graphemes be identified and adapted in the revision of the vocalic graphemes.

In identifying, discussing and accounting for discrepancies between the phonemic and the graphemic structure of Gĩkũyũ, this study makes a contibution to the efforts that are currently being made to harmonize the orthography of Bantu

languages in Kenya. A similar study to this one should be done for the other Bantu languages an an endeavour to inform these efforts.

This study has shown that the tenets of OT are applicable to orthography. In the development and revision of orthographies, these tenets should be applied with the anticipation that the pitfalls observed by this study could be avoided in any other orthography.

REFERENCES

Anthony, L. (2013). A critical look at software tools in corpus linguistics. *Linguistic Research 30(2), 141-161.*

Armstrong, L.E. (1967). *The Phonetic and Tonal Structure of Kikuyu*. London: Dawsons of Pall Mall.

Baker, P. (1997). Developing Ways of Writing Vernaculars: Problems and Solutions in a Historical Perspective. In Andrée Tabouret-Keller, Robert B. Le Page, PenelopeGardner-Chloros, and Gabrielle Varro (eds.), *Vernacular Literacy: A Re-Evaluation*. Oxford: Clarendon Press, 93–141.

Baker, R. G. (1980). Orthographic Awareness. In Uta Frith *Cognitive Processes in Spelling.* London: Academic Press Inc.

Banda, F. (ed) (2002). *Language across Border*. Cape Town: CASAS.

Barnwell, K. (1998). *Course Notes of the Introductory Course in Applied Linguistics*. Jos. Nigeria Bible Translation Trust.

Baroni, A. (2016) *Constraint Interaction and Writing Systems Typology*. Dossiers d'hel, SHESL, 2016, Ecriture(s) et representations du langage et des langues, 9, pp. 290-303. <http://htl.linguist.univparisdiderot.fr/hel/dossiers/numero9>.<hal013053>

Benson, T. G. (1964). *Kikuyu – English Dictionary.* Oxford: Oxford University Press.

Boersma, P. (1998). *Functional Phonology: Formalizing the interactions between articulatory and perceptual drives*. PhD Dissertation. University of

Amsterdam.UvA-DARE, the Institutional Repository of the University of Amsterdam (UvA) http://dare.uva.nl/cgi/arno/show.cgi

Boersma, P. (2014). *Functional phonology: Formalizing the Interaction between Articulatory and Perceptual drives.* Downloaded from UvA-DARE, the Institutional Repository of the University of Amsterdam (UvA) http://dare.uva.nl/cgi/arno/show.cgi

Bradley, D. (2003). *Issues in Orthography Development and Reform*. La Trobe University, Australia.

Bradley, D. (2002). Language attitudes: The key factor in language maintenance. In Bradley D. & Maya (eds). *Language endangerment and language maintenance*.

Cahill, M. and Karan, E. (2008). *Factors in designing effective orthographies for unwritten languages*. SIL Electronic Working Paper 2008-001. http://www.sil.org/silewp/2008/silewp2008-001.pdf

Casali, F. R. (2016). Some Inventory-Related Asymmetries in the Patterning of Tongue Root Harmony Systems. In *Studies in African Linguistics* Vol. 45, Numbers 1&2, 2016.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English.* New York: Harper & Row.

Choti, J. (2015). Phonological Asymetries of Bantu Nasal Prefixes. In Kramer, R., Zsiga, E. C. & Boyer, O. T. (Eds,). *Selected Proceedings of the 44th Annual Conference on African Linguistics* (pp. 37-51). Somerville, MA.

Clements, G.N. (1991). Vowel Height Assimilation in Bantu Languages. In *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society: Special Session on African Language Structures,* (pp. 25-64).

Coulmas, F. (1989). *The Writing Systems of the World*. Blackwell: Oxford.

Coupe, A. R. (2014). Strategies for Analyzing Tone Languages. In *Language Documentation and Conservation (8),* (pp. 462-489*).*

Farris-Trimble, A. W. (2008). *Cumulative Faithfulness Effects in Phonology*. Unpublished PhD Thesis, Indiana University

Flemming, E. (2002). *Auditory Representations in Phonology.* New York: Routledge.

Fromkin, V., Rodman, R. & Hyams, N. (2007). *An Introduction to Language.* Boston: Thomson Wadsworth.

Frost, R. (1989). Orthography and Phonology: The Psychological Reality of Orthographic Depth. In *Haskins Laboratories Status Report on Speech Research 1989,* S-R-99/100, 162-171

Githiora, C. (2003). Gĩkũyũ orthography: Past and Future Horizons. In Githiora, C., Littlefield, H. & Manfredi, V. (Eds). *Kinyira njira! Step Firmly on the Pathway! Trends in African Linguistics 5* (85-95, Lawrenceville, New Jersey*:* Africa World Press.

Githiora, C. (2016). *Gĩkũyũ*. London: Elsevier.

Goldsmith, J. (1990). *Autosegmental and Metrical Phonology.* London: Basil Blackwell.

Guérin, V. (2008). Writing an Endangered Language. In *Language Documentation & Conservation Vol. 2(1).*

Guthrie, E. (1967). *Classification of Bantu Languages*. London: Dawsons of Pall Mall.

Hasselbring, S. A. (2006). *Cross Dialectal Acceptance of Written Standards*. Unpublished PhD Thesis, University of South Africa.

Hayes, B. P. (1980). *A Metrical Theory of Stress Rules*. MIT.

Iribemwangi, P. I. (2012). A Case for the Harmonization of Kikuyu, Kiembu and Kimbeere Phonology and Orthography. In Ogechi, N. A., Oduor, J. A. N. & Iribemwangi (eds.) *The Harmonization and Standardization of Kenyan Languages*: *Orthography and Other Aspects* (pp.20-38). Cape Town: CASAS.

Jeon, L., Li, J., Mauney, S., Navarro, A. & Wittke, J. (2015). A Basic Sketch Grammar of Gĩkũyũ. In Englebretson, R. (ed). *A Special Issue of Rice Working Papers in LinguisticsVolume VI.*

Kager, R. (1999). *Optimality Theory*. New York: Cambridge University Press.

Kang, H. & Kos, S. (2011). A Phonetic Study of the Tongue Root Contrast in Buriat and Ewen. In *LSA Annual Meeting, 7 January, 2011,* Pittsburg. USA.

Karan, E. (2006). *Writing System Development and Reform: A Process.* Unpublished MA Thesis, University of North Dakota.

Katz, L. & Frost, R. (1992). *Orthography, Phonology, Morphology and Meaning: An Overview*. Elsevier Science Publisher

Kenya National Bureau of Statistics. (2010). *The* 2009 *Kenya Population and Housing Census: Population Distribution by Age, Sex, and Administrative Units, Vol. 1c.* Nairobi: Government Press.

Kerr, H. (1998). *The Cognitive Psychology of Literacy Teaching: Reading, Writing, Spelling, Dyslexia (& a bit besides).* www.hugokerr.info

Kioko, A. N. (2002). A Case for Improved Orthography. In Owino, F. R. (ed.) *Speaking African: African Languages for Education and Development*, (pp. 231-241). Cape Town: CASAS.

Kioko, A. N. (2005). *Theoretical Issues in the Grammar of Kikamba: A Bantu Language*. Muchen: Lincom Europa.

Kioko, A. N., Njoroge, M. C., Kuria, P. M. (2012). Harmonizing the Orthography of Gĩkũyũ and Kĩkamba. In Ogechi, N. A., Oduor, J. A. N. & Iribemwangi (eds.) *The Harmonization of and Standardization of Kenyan Languages*: *Orthography and Other Aspects* (pp. 39-63). Cape Town: CASAS.

Kioko, A. N., Mbori, B., Kitetu, C., Ongarora, D., Buliba, F. A., Mworia, F., Chai, F., Basuba, H., Mwaliwa, H., Chacha, L., Saidi, M. V., Njoroge, M., Lonyangapuo, M., Njoka, M., Ndung'u, R. & Ogechi, N. (2012). *A United Orthography for Bantu Languages of Kenya*.Cape Town: CASAS.

Kuria, P.M. (2005). *An Examination of the Discrepancies between the Graphemic and the Phonological Structure of Gĩkũyũ*. Unpublished M.A. Thesis, Kenyatta University.

Labov, W. (1991). The intersection of sex and social class in the course of linguistic change. In *Language Variation and Change,* 2 (pp. 205-254). Cambridge University Press.

Massamba, D. P. M. (1996). *Phonological Theory: History and Development.* Dar-es-Salaam: Dar-es-Salaam University Press.

Malone, S. (2004). *Manual for developing Literacy and Adult Education Programs in Minority Language Communities*. Bangkok: UNESCO. http://unesdoc.unesco.org/images/0013/001351/135164e.pdf23.5.2012 5.01pm

Mason, J. (2002). *Qualitative Researching*. London: SAGE Publications Ltd.

McCarthy, J. and Prince A. (1995). "Faithfulness and Reduplicative Identity." In *University of MassachusettsOccasional Papers in Linguistics 18: Papers in Optimality Theory*. GLSA, University ofMassachusetts, Amherst. [ROA-60.]

McCarthy, J. J. (2007). *What is Optimality Theory?* University of Massachusetts, Amherst.

Meyerhoff, M. (2006). *Introducing Sociolinguistics*. London: Routledge.

Ministry of Education (2002). *Primary Education Syllabus Vol. 1*. Nairobi: Kenya Literature Bureau

Mutahi, E. K. (1977). *Sound Change and the Classification of the Dialects of Southern Mt. Kenya.* Unpublished PhD Thesis, University of Nairobi.

Mwangi, P. (2001). *Verb Morphology in Gĩkũyũ in the Light of Morphosyntatic Theories* Unpublished PhD Thesis, Kenyatta University.

Mwangi, P.W., Njoroge, M. C and Mose, E. G. (2013). Harmonizing the Orthographies of Bantu Languages: The case of Gĩkũyũ and Ekegusii in Kenya. In *The University of Nairobi Journal of Language and Linguistics, Vol.3* (pp. 108 – 122*).*

Mwihaki, A. N. (1998). *Loanword Nativization: A Generative View of the PhonologicalAdaptation of Gĩkũyũ Loanwords*. Unpublished PhD Thesis, Kenyatta University.

Nabea, W. (2009). Language Policy in Kenya: Negotiation with Hegemony In *The Journal of Pan African Studies*, Vol.3, (1) (pp. 121-138).

Ngure, K. K. (2005). *The Loss of Prenasalisation in the Northern and Southern Varieties of Gĩkũyũ* . Unpublished M.A. Thesis, University of Nairobi.

Okombo, O. (2010). Kenya's Proposed Devolution of Governance Structures: Its Implications for Language Policy and Nationhood. *International Conference on Multilingualism 15.*

Omachonu, G.S. (2008). Comparative Optimality Theory Analysis of Primary Stress Assignment in Standard English and Nigerian English. In *Journal of Language and Translations 9-1*.

Padgett, J. (1979). Perceptual Distance of Contrast: Vowel Height and Nasality. In Karvonen, D., Katayama, M. & Walker, R. (eds). *Phonology at Santa Cruz (PASC)* (pp 63-78). Santa Cruz, C.A..

Padgett, J. (2003). Contrast and Post-velar Fronting in Russian. *Natural Language and Linguistic Theory,21* (pp. 39 - 87).

Peng, L. (1991). Swati and Kikuyu Reduplication: Evidence against Exhaustive Copy. In *Studies in African Linguistics Volume* 22 (pp 45-72).

Peng, L. (2000). Kikuyu Vowel Harmony. In *South African Journal of African Languages, 20(4).*

Peng, L. (2003). Local Conjunction and Kikuyu Consonant Mutation. In *Proceedings of the 26th Annual Penn.* University of Pennsylvania Working Papers in Linguistics Vol. 9 Issue 1.

Perry, C., Ziegler, J. C., & Colheart, M. (2002). How Predictable is Spelling? Developing and Testing Metrics of Phoneme – Grapheme Contingency. In *The Quarterly Journal of Experimental Psychology*, 2002, 55A(3) (pp. 897-915)

Prah, K. (2003). "Harmonising and Standardizing African Languages for Scientific and Technological Development: The CASAS experience. Paper presented to *The Icelandic Language Council: Conference on Language Policy and Standardization*. Reykjavik, Iceland.

Prince, A.S. and Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Ms., Rutgers University, Rutgers, N.J. and University of Colorado at Boulder. Boulder, Col.

Republic of Kenya (1976). *Report of the National Committee on Educational Objectives and Policies (Gachathi Report).* Nairobi: Government Printers.

Republic of Kenya (2017). *Basic Education Curriculum Framework.* Nairobi: Kenya Institute of Curriculum Development.

Republic of Kenya (2019). *2019 Kenya Population and Housing Census: Vol I & II.* Kenya National Bureau of Statistics.

Ruddell, S. (2013). *An acoustic study of the Japanese voiceless bilabial fricative.* San Francisco State University.

Sampson, G. (1985). *Writing Systems*. London: Hutchison.

Sampson, G. (2016). Writing Systems: Methods for Recording Language. In Keith Allan (ed.). *Routledge Handbook of Linguistics.* New York: Routledge.

Sasaki, M. (2003). The Writing System of an Artificial Language: For Efficient Orthographic Processing. In *Journal of Universal Language* 4, (pp. 91-112).

Schroeder, L. (2008). Bantu Orthography Manual. *SIL International*: SIL e-books.

Seifart, F. (2006). Orthography Development. In Gippert, J., Himmelmann N., & Ulrike Mosel U (Eds), *Essentials of language documentation* (pp. 275–299) New York: Mouton de Gruyter.

Smalley, W. A. (1964). How shall I Write this Language? In *Orthography Studies: New Articles on New Writing System.* London: Oxford University Press.

Smith, E.J. M. (2004). Using Optimality Theory to Reconstruct Elamite Phonology. *LIN 1290 Forum Paper.*

Stark, J. E. (2010). Kamburi Orthography Design. *SIL International*.

Starwalt, C. G. A. (2008). *The Acoustic Correlates of ATR Harmony in Seven- And Nine Vowel African Languages: A Phonetic Inquiry into Phonological Structure*. Unpublished PhD Thesis, the University of Texas at Arlington.

Stegen, O. (2005). Vowel Length in Eastern Bantu Orthographies. *SIL International.*

van der Hulst, H. (2014). *Lowering Harmony in Bantu: An RcvP Account*. John Benjamin Publishing Co.

van Oostendorp, Marc. (2004). *The Theory of Faithfulness*. Amsterdam: Meertens Institute.

Venezky, R. L. (2004). In Search of the Perfect Orthography. In *Written Language and Literacy 7:2* (139-163). John Benjamins Publishing Co.

Wangia, J. I. (2003). *Aspects of Mistranslation in the 1951 Lulogooli Bible*. Unpublished PhD Thesis. Kenyatta University.

Wanjau, G. (1991). *Mwandikire wa GiGĩkũyũ Karin'ga: Ibuku ria Keeri*. Karatina: Gakaara Press Ltd.

Wanjau, G. (1989). *Thooma GĩGĩkũyũ Kĩega: Ibuku rĩa Mbere.* Karatina: Gakaara Press Ltd.

Wardhaugh, R. (2006). *An Introduction to Sociolinguistics*. Oxford: Blackwell.

wa Thiong'o, N. (2004). *Murogi wa Kagogo*. Nairobi: East African Educational Publishers Ltd.

Waweru, M. M. (2011). *Gĩkũyũ Verbal Extensions: A Minimalist Analysis*. Unpublished Phd Thesis, Kenyatta University.

Williamson, K. (1984). *Practical Orthography in Nigeria*. Ibadan: Heinman

Wiese, R. (2004). How to Optimize Orthography. In *Written Language and Literacy 7:2* (pp. 305-331). John Benjamins Publishing Co.

Wiese, R. and Song, H. J. (2010) Resistance to complexity interacting with visual shape-German and Korean orthography. In *Writing Systems Research, Vol. 2, No. 2* (pp. 87–103). Oxford University Press.

Appendix 1

QUESTIONNAIRE

Age: County: Male/Female (tick one)

SECTION ONE: Tick the response you think is most correct for you.

1. Writing in Gĩkũyũ is easy.

a) I strongly agree b) I agree c) I disagree d) I strongly disagree

2. Reading in Gĩkũyũ is easy.

a) I strongly agree b) I agree c) I disagree d) I strongly disagree

3. Reading and writing in English is easier than doing so in Gĩkũyũ.

a) I strongly agree b) I agree c) I disagree d) I strongly disagree

4. Tick the letter or letters that are most difficult for you in reading Gĩkũyũ.

a, e, i, ĩ, o, u, ũ.

Give a reason(s) why this is so.

5. Tick the letter or letters you find most confusing when reading.

m, b,t, th, n, ng', k, ny, h,c, y, w, mb, nd, nj, ng.

Give a reason(s) why this is so.

6. Which letter or letters do you think should not be used in writing Gĩkũyũ?

Why?

7. Apart from the letters in No. 6 above, which others would you suggest that they be used in writing Gĩkũyũ?

Why?

SECTION TWO: Translation

Read the following sentences. Translate the underlined words into Gĩkũyũ (give one word only, not a phrase). If two translations are possible, give them.

1. The bus comes at 1.00 pm.

   ____ _____

2. The small boy put the oranges in the drawer.

   _____ _____ _____

3. The judge bought four litres of petrol.

   _____ ____ _____ _____

4. I don't want questions, actions only.

   _____ _____ _____

5. His dream is to own a new shop.

   _____ ____ _____

6. We like our cow and all dogs.

   ___ ___ ___ _____

7. This ugly cat is theirs.

   ____ ____ _____

8. The snake will swallow the buffalo.

   ____ _____ _____

9. The middle-aged woman saw two trees in the garden.

   _____ _____ ___ ____ _____

10. The driver gives the young men pangas.

    _____ _____ _____ _____

11. God give us your peace, we pray.

    ___ _____ ___ ____ ____

12. The food is in the kitchen near the lamp.

    _____ _____ _____

13. All the people saw the bats at the mill.

    ___ ____ _____ _____

14. A small he-goat, grey in colour jumped across the road.

_____ _____

15. Do you know those two lazy people who are yawning?

_____ _____ _____

16. Follow your parent's advice: go tomorrow.

_____ ___ _____

17. The acacia tree burnt for four hours.

_____ _____ _____

18. He has a lot of maize stalks inside the barn.

_____ _____

19. Leave the house! He ordered the children.

_____ _____

20. He has lost money through corruption.

_____ _____ _____

21. He has said that you help them to go there with the towel and baby shawl.

_____ _____ _____ ____ _____

22. Both of you put clothes in the latrine that was built many years ago.

_____ _____ _____ _____ _____

23. Interpret: the drunkard cut his arm with a razor blade.

_____ _____ _____ _____

24. We look like flowers with our eyes shaded.

___ _____ _____ _____

25. He died a terrible death after they cut his hair with a traditional blade.

_____ _____ _____ _____

26. Carry the gift and I carry the peeled peas and the cooking stick.

____ ____ _____ _____ _____

27. She knows his words betray who he married two months ago.

_____ _____ _____ _____

28. He spoke plainly. That's why she gets water from the two rivers.

_____ _____ _____ _____

29. Eat all that food and you will belch all night!

____ _____

30. Share out the food to your co-wives. Let them also eat.

_____ _____ ____ _____

31. Envy and loitering are not known for nation building.

____ _____ _____ _____

32. Relieve Njeri of her load and stop admiring her thighs!

_____ _____

33. Give her her inheritance though she doesn't know why.

_____ _____

34. They tried to cut the dog's tail but were stopped by their parents.

_____ _____ _____

35. Let me search for a job. Pray for me.

_____ _____

SECTION III: Reading

Kindly read the words you have translated. You will be taped as you read the words.

Appendix II

WORDS WITH TARGET GRAPHEMES IN RESPECTIVE POSITIONS

CONSONANTAL GRAPHEMES

Phoneme/ grapheme		Word initial position	Gloss	Word medial position	Gloss
/^{m}b/	<mb>	**mb**athi	bus	ng'o**mb**e	Cow
/β/	<b>	**b**etũrũ	petrol	ma**b**anga	machetes
/m/	<m>	**m**ũgwanja,	seven	kũ**m**eria	swallow
/ɲ/	<ny>	**ny**au	cat	i**ny**a	Four
/t/	<t>	**t**ũhe	give us	ri**t**a	Litres
/ð/	<th>	**th**ayũ	peace	mba**th**i	Bus
/^{n}d/	<nd>	**nd**uka	shop	mũgũ**nd**a	garden
/r/	<r>	**r**ita	litres	ĩĩ**r**i	Two
/n/	<n>	**n**ĩũĩ	do you know	a**n**ake	young men
/ʃ/	<c>	**c**iũria	questions	ma**c**ungwa	oranges
/^{ɲ}ɟ /	<nj>	**nj**anji	judge	mũgwa**nj**a	Seven
/k/	<k>	**k**ĩrooto	dream	nyo**k**a	Snake
/^{n}g/	<ng>	**ng**ai	God	uu**ng**umania	corruption
/ŋ/	<ng'>	**ng'**ombe	cows	njo**ng'**i	Ugly
/w/	<w>	**w**aku	yours	ta**w**a	Lamp
/ h/	<h>	**h**uuhu	owls	tũ**h**e	give us
/j/	<y>	**y**ao	theirs	tha**y**ũ	Peace
/ɣ/	<g>	gathenge	small he-goat	mbo**g**o	buffalo

VOCALIC GRAPHEMES

Phoneme/ grapheme		Word initial position	Gloss	Word medial position	Gloss	Word final position	Gloss
/a/	<a>	**a**gũũta	lazy people	y**a**o	theirs	ndereb**a**	Driver
/ɛ /	<e>	**e**rĩ	two people	nd**e**reba	drive	onir**e**	s/he saw
/i/	<i>	**i**rio	Food	ir**i**o	food	njong'**i**	Ugly
/e/	<ĩ>	**ĩ**kumeria	will swallow	k**ĩ**roto	dream	ĩĩr**ĩ**	Two
/ɔ/	<o>	**o**nire	s/he saw	ny**o**ka	snake	ya**o**	Theirs
/u/	<u>	**u**ma	move out	nd**u**ka	shop	nya**u**	Cat

/o/ <ũ>	**ũ**mũthĩ	Today	bet**ũ**rũ	petrol	thay**ũ**	Peace

LONG VOWELS

Phoneme/ grapheme	Word initial position	gloss	Word medial position	gloss	Word final position	Gloss
/a:/ <aa>	**a**nake	young men	mat**aa**ro	advice	mug**aa**	Acacia
/ɛ:/ <ee>	**e**rĩ	two (people)	mab**ee**be	maize stalks	at**ee**	s/he lost
/i:/ <ii>	**i**tu	ours	r**ii**ko	kitchen	munab**ii**	Prophet
/e:/ <ĩĩ>	**ĩ**rĩ	two	n**ĩĩ**kameria	will swallow	kah**ĩĩ**	small boy
/ɔ:/ <oo>	**o**the	all	kĩr**oo**to	dream	ndir**oo**	Drawer
/u:/ <uu>	**u**ngumania	corruption	h**uu**hu	bats	kab**uu**	Grey
/o:/ <ũũ>	**ũ**ĩ	s/he knows	ag**ũũ**ta	lazy people		

Concatenated Vowels

Phoneme/ Grapheme	Initial	gloss	Medial	gloss	Final	gloss
/ie/ <iĩ>	-		c**iĩ**ko	actions	muc**iĩ**	home
/iɛ/ <ie>	-		c**ie**ro	thighs	njar**ie**	i search
/ia/ <ia>	-		c**ia**na	children	ungumani**a**	corruption
/iɔ/ <io>	-		c**io**the	all	irio	food
/io/ <iũ>	-		c**i**ũria	questions	rũc**iũ**	tomorrow
/iu/ <iu>	-		c**iu**go	words	-	
/ei/ <ĩi>	-		th**iĩ**nĩ	inside	gĩth**ĩi**	mill
/eu/ <ĩu>	-		-		mũr**ĩu**	drunkard
/eo/ <ĩũ>	-		k**ĩũ**ria	question	-	
/eɛ/ <ĩe>	-		m**ĩe**ri	months	ar**ĩe**	to let eat
/eɔ/ <ĩo>	-		k**ĩo**ro	latrine	nĩk**ĩo**	that's why
/ea/ <ĩa>	-		m**ĩa**ka	years	R**ĩa**	eat
/ɛi/ <ei>	-		t**ei**thia	help	kang'**ei**	middle-aged woman
/ɛɔ/ <eo>	-		-		kĩh**eo**	gift
/ɛo/ <eũ>	-		-		m**ĩeũ**	yawn
/ɛa/ <ea>			ah**ea**ga	gives	theb**ea**	belch
/ai/ <ai>	**ai**ru	co-wives	m**ai**tho	eyes	ig**ai**	share
/ae/ <aĩ>	-		g**aĩ**ra	share out	m**aĩ**	water
/aɔ/ <ao>	-		-		w**ao**	theirs
/ao/ <aũ>	aũra	relieve	t**aũ**ra	interpret	git**aũ**	speak

						plainly
/au/ <au>	-		t**au**rũ	towel	ny**au**	cat
/ɔi/ <oi>	**oi**ga	says	tor**oi**ra	loiter	ng**oi**	baby carrier
/ɔe/ <oĩ>	**oĩ**	knows	nd**oĩ**kaine	is unknown	nd**oĩ**	doesn't know
/ɔɛ/ <oe>	-		h**oe**ra	pray for	mũtong'**oe**	tail
/ɔa/ <oa>	-		-		ng'**oa**	assemble
/ou/ <ũu>			-		k**ũu**	there
/oi/ <ũi>	**ũi**ru	jealosy	m**ũi**ko	cooking stick	kany**ũi**	razor blade
/oe/ <ũĩ>	ũĩkaine	known	g**ũĩ**ta	call	nj**ũĩ**	rivers
/oɛ/ <ũɛ>	-		r**ũe**nji	traditional razor	nj**ũe**	peeled
/oa/ <ũa>	-		g**ũa**ka	to build	ih**ũa**	flower
/ui/ <ui>	-		iny**ui**ki	you (pl) alone	ng**ui**	dog
/ue/ <uĩ>	-		nj**uĩ**rĩ	hair	ith**uĩ**	we
/uɛ/ <ue>	-		iny**ue**rĩ	both of you	ng**ue**	i die
/ua/ <ua>	-		h**ua**na	look like	k**ua**	die
/uɔ/ <uo>	-		g**uo**ko	arm	ng**uo**	dress
/uo/ <uũ>	-		-		gĩk**uũ**	death

Appendix III

KENYATTA UNIVERSITY GRADUATE SCHOOL RESEARCH AUTHORIZATION

KENYATTA UNIVERSITY
GRADUATE SCHOOL

E-mail: dean-graduate@ku.ac.ke
Website: www.ku.ac.ke

P.O. Box 43844, 00100
NAIROBI, KENYA
Tel. 8710901 Ext. 57530

OUR REF: C82/11301/08

Date: 2nd February, 2018

The Director General
National Commission for Science,
Technology & Innovation
P.O. Box 30623-00100,
NAIROBI

Dear Sir/Madam,

RE: RESEARCH AUTHORIZATION FOR MR. KURIA P. MBURU REG. NO. C82/11301/08

I write to introduce **Mr. Mburu** who is a Postgraduate Student of this University. He is registered for Ph.D. Degree programme in the **Department of English & Linguistics in the School of Humanities & Social Sciences.**

Mr. Mburu intends to conduct research for a Ph.D. thesis entitled, **"The Gikuyu Graphemic Structure: An Optimality Approach"**

Any assistance given will be highly appreciated.

Yours faithfully,

MRS. LUCY N. MBAABU
FOR: DEAN, GRADUATE SCHOOL

HI/cao

Appendix IV

NACOSTI PERMIT

Permit No : NACOSTI/P/18/95240/21525
Date Of Issue : 2nd March,2018
Fee Recieved :Ksh 2000

THIS IS TO CERTIFY THAT:
MR. PETER MBURU KURIA
of KENYATTA UNIVERSITY, 0-902
KIKUYU,has been permitted to conduct
research in *Kiambu , Kirinyaga ,*
Muranga , Nyeri Counties

on the topic: *GIKUYU GRAPHEMIC*
STRUCTURE: AN OPTIMALITY APPROACH

for the period ending:
2nd March,2019

.....
Applicant's
Signature

.....
Director General
National Commission for Science,
Technology & Innovation