

*To My parents –
Shyamkishore & Khambiton*

With love

Krishna

DECLARATION

I, Ksh. Krishna Bati Singha, bearing Registration No. Ph.D/1166/2010, dated 15/09/2010, hereby declare that the subject matter of the thesis entitled “**MANIPURI MORPHOLOGICAL ANALYSIS WITH FINITE STATE TECHNIQUES**” is the record of work done by me and that the contents of this thesis did not form the basis of reward of any degree to me or anybody else to the best of my knowledge. The thesis has not been submitted in any other University/ institute.

Place: Silchar

Date: 12/5/2014

(Ksh. Krishna Bati Singha)

ACKNOWLEDGEMENTS

It is a wonderful feeling that some of my “hobbies” during the earlier phase of my life materialize in a doctoral thesis. I used to take interest in exploring the similarity and differences among different languages. A question was always in my mind that why languages differ in their ways of expression (words, sentences, etc.) while they would mean the same thing. Little did I know about morphology or morphemes at that time. Fascinating thing about computers these days is that computers act intelligently and skillfully. It has reduced many of our day to day workloads. Human nature is greedy and they want more with every achievement. And it succeeds also in many fields. How about making the computer understand and interact in natural language? Simply, make a mathematical model for natural language because it understands only mathematical interpretations. With this research study I came to know about the beauty of Mathematics.

Developing a mathematical model for a natural language was not an easy task for me. Many people have contributed in their own ways to help complete my thesis work. My supervisor, Dr. Bipul Syam Purkaystha, suggested me the different ways of processing morphological analysis. With his consistent push towards a mathematically more sound and elegant model formulation made me land here with this outcome. Thank you sir!

I profoundly thank all my research mates in the department of Computer Science for their valuable support, with whom I used to discuss and share, turn to, at my difficult days of the research work. The office staff of the department and the university as a whole helped me in official matters whenever required and am very thankful to them. I also take this opportunity to show my gratitude and thank Associate Professor, Dr. Kh. Dhiren Singha, Department of Linguistics and Professor Dr. W. Raghmani Singh, Department of Manipuri, Assam University, Silchar, for their extensive valuable suggestions and information in clearing my doubts about the linguistics aspects of Manipuri language.

I am grateful to all the anonymous reviewers who have contributed to the quality of the publications through their insightful comments.

Without my mother, I would have never dared to do this study. I want to thank my mom for the uncountable amount of hours of babysitting, which allowed me to complete my thesis.

My father and brothers played a key role in supporting and helping me out whenever their help is required along the course of my journey to complete the thesis.

Last but not the least, I want to say sorry to my spouse and daughter Trisha for the time I spent on my research work, that would otherwise have been for them, and thank them for their support and understanding all the way to complete the study.

Thank you everyone!

Date: 12/5/2014

Place: Silchar

Ksh. Krishna Bati Singha

CONTENTS

LIST OF TABLES	IX
LIST OF FIGURES.....	X
ABSTRACT.....	XI
CHAPTER 1 : INTRODUCTION	1
1.1 Morphological Analysis.....	2
1.2 Motivation.....	4
1.3 Objectives.....	5
1.4 Outline of the thesis	6
CHAPTER 2 : REVIEW OF LITERATURE.....	8
2.1 Methodology	13
2.2 Design issues	14
2.3 Limitations	14
CHAPTER 3 : MANIPURI	15
3.1 Origin and information of Manipuri Language.....	16
3.2 Linguistic features	17
3.3 Writing system and sound system	18
CHAPTER 4 : FINITE STATE MORPHOLOGY- THE THEORETICAL FRAMEWORK.....	22
4.1 Origin	23
4.2 Finite-State Technology.....	23
4.3 Regular Expression (RE) vis-a-vis Finite-State Automata	26
4.4 Regular Languages and Regular Relation.....	30
4.5 Finite-State Transducer.....	33
4.5.1 Some Important Operations on Finite-State Transducers.....	35
4.5.2 Closure properties of Finite-State Transducers.....	39
4.6 Two-level Morphology	41
4.7 Xerox finite-state tool	42
4.7.1 lexc	43
4.7.2 xfst.....	44
CHAPTER 5 : MANIPURI MORPHOLOGY	46
5.1 Manipuri morphemes.....	48

5.1.1 Roots.....	49
5.1.2 Affixes- Suffixes and Prefixes	50
5.2 Word classes.....	56
5.3 Manipuri morphotactics/ morphosyntax	69
5.4 Manipuri morphophonological variation/ alternation rules	78
5.5 Morphological Ambiguity.....	84
5.6 Linguistic insight	85
CHAPTER 6 :FINITE STATE MORPHOLOGICAL ANALYSIS WITH XFST	
TOOL.....	86
6.1 Manipuri source lexicon.....	87
6.2 The Rule Transducer.....	93
6.3 Applying Composition operation	95
6.4 The Lexical Transducer	97
CHAPTER 7 : EVALUATION AND ANALYSIS	
7.1 Evaluation	98
7.2 Efficiency and sufficiency.....	101
7.3 Coverage	102
CHAPTER 8 :CONCLUSION.....	
Resolved research questions	103
Contribution of the research wok	104
Open research question(s).....	104
Future directions of research.....	105
REFERENCES.....	106
ANNEXURE- I : ABBREVIATIONS	111
ANNEXURE II: MULTICHAR SYMBOLS	113
ANNEXURE III: LANGUAGE MODEL FOR ADJECTIVES.....	115
ANNEXURE IV: MODEL OF 5 WH-WORDS.....	118
ANNEXURE V: PAPERS PUBLISHED.....	121
ANNEXURE VI: CONFERENCES AND WORKSHOPS ATTENDED.....	122

LIST OF TABLES

Table 3-1: Bengali vowel symbols used in written Meiteiron	18
Table 3-2: Bengali consonant symbols used in written Manipuri.....	19
Table 3-3: Manipuri consonants and corresponding Bengali letters.....	19
Table 3-4: Manipuri monophthongs and corresponding Bengali letters	20
Table 3-5: Meiteiron diphthongs and corresponding Bengali letters	20
Table 3-6: Dependent vowel signs (Matra/ Cheitap)	21
Table 4-1: Notations For Regular Expression Basic Symbols.....	27
Table 4-2: Notations For Multi-Character Symbols.....	27
Table 4-3: Special symbol-like notations	28
Table 4-4: Grouping, dteration and optionality operators	28
Table 4-5: Operations on REs	28
Table 4-6: Examples of RE patterns and their meaning	29
Table 4-7: Example of an RE with corresponding Regular Language.....	31
Table 4-8: Operations on RE and RL	32
Table 5-1: Morphemes with lexical category	47
Table 5-2: Pronominal prefixes.....	51
Table 5-3: List of some non-pronominal prefixes.....	52
Table 5-4: Nominal suffixes	54
Table 5-5: Verbal suffixes	55
Table 5-6: Free animate nouns with some inflected forms.....	57
Table 5-7: Bound roots from kinship terms with pronominal markers	59
Table 5-8: Free and bound forms of Personal Pronouns	60
Table 5-9: Idiosyncratic features of attaching suffixes to Personal Pronouns.....	61
Table 5-10: Some wh-words with derived forms.....	62
Table 5-11: Affixes used in adjective formation.....	66
Table 5-12: Morphotactics Of Verb Suffixes	70
Table 5-13: Morphotactics of nominal suffixes	76
Table 6-1: A script file to compose the adjective lexicon with rules	95
Table 7-1: Statistics of the lexical entry and total word forms per entry	99
Table 7-2: Analysis of nominal category word forms.....	99
Table 7-3: Analysis of verbal category word forms.....	100

LIST OF FIGURES

Figure 3-1: Schematic diagram of Sino-Tibetan (Adapted From Http://Stedt.Berkeley.Edu/Html/Stfamily.Html)	16
Figure 4-1: A Finite-State Automaton.....	24
Figure 4-2: Finite-State Transition Network For RE [a b* c].....	30
Figure 4-3: FST for some Bengali Vowel letters to its corresponding matra.....	34
Figure 4-4: FST for noun stems	35
Figure 4-5: FST for adverbs.....	35
Figure 4-6: FST for sdjectives.....	35
Figure 4-7: Union of noun, adverb and adjective FSTs	36
Figure 4-8: FST for noun stem "ঈশিং"/ water	36
Figure 4-9: FST for copula নি, genitive গী, and accusative suffix বু.....	36
Figure 4-10: Concatenated FST for the above two FSTs	37
Figure 4-11: FSTs for <অচৌব+Adj+PI, অচৌবশিং> and <অচৌবশিং, ঞcoubəsiŋ>	37
Figure 4-12: FST after applying composition on the above two FST networks.....	38
Figure 4-13: Block Diagram of Transducer composition.....	41
Figure 4-14: Example showing Transducer composition.....	41
Figure 5-1: Types of morphemes (Source: Ch. Yashwanta Singh, 2000).....	48
Figure 5-2: Types of nouns	57
Figure 5-3: FST for free animate nouns with gender and some case markers.....	58
Figure 5-4: FST for kinship terms with pronominal markers and some case markers	59
Figure 5-5: FST for free and bound form Pronouns.....	60
Figure 5-6 FST for Demonstrative Pronoun	61
Figure 5-7 FST for adjective formation.....	67
Figure 5-8 FST for analysis of adjective <ইরিব>	68
Figure 5-9: An FST for aspect suffix	76
Figure 6-1: The Lexical Transducer	97