GRAPHEME-TO-PHONEME CONVERSION FOR FOREIGN WORDS IN MYANMAR LANGUAGE

SWE ZIN AUNG

M.C.Sc. **SEPTEMBER 2022**

GRAPHEME-TO-PHONEME CONVERSION FOR FOREIGN WORDS IN MYANMAR LANGUAGE

By

SWE ZIN AUNG

B.C.Sc

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Computer Science

(M.C.Sc.)

University of Computer Studies, Yangon

September 2022

ACKNOWLEDGEMENTS

I would like to thank Ministry of Education and my Principle at No.177 Basic Education Middle School and No.8 Basic Education High School(Pathien) for providing and permission on duty during the Master course work and Thesis research at the University of Computer Study.

Firstly, I would like to express my sincere gratitude to **Dr. Mie Mie Khin**, Rector of the University of Computer Studies, Yangon, for her kind permission to submit this research.

I would like to express gratitude and deepest appreciation my supervisor, **Dr. Aye Mya Hlaing**, Associate Professor, NLP Lab, University of Computer Studies, Yangon, for providing some useful related documents, her kindly suggestions and encouragement, excellent mentorship, invaluable guidance, constructive comment, critical thinking and practical advice during these research completion.

I sincerely wish to thank **Dr. Khin Mar Soe** and **Dr.Win Pa Pa,** Professors of the Natural Language Processing Lab, University of Computer Studies, Yangon for providing the necessary support for research and permission to research at Natural Language Processing Lab, University of Computer Studies, Yangon during the preparation of thesis research.

I also like to express to special thanks to our deans **Dr. Si Si Mar Win** and **Dr. Tin Zar Thaw**, Professors of the Faculty of Computer Science, for their communication to do my research and effectively guidelines during research completion study.

I would like to extend deeply and special thanks to **Daw Hnin Yee Aung,** Lecturer, Department of English, for her support in writing my research papers and my dissertation from the language point of view. I also like to express special thanks for suggestions and recommendations of the teachers who attended all my seminars in their free time.

I would like to extend my deepest gratitude to my warden, **Dr. Win Lae Lae Phyu,** Professor, Faculty of Computer Science for permission to stay Ziwaka Hostel, University of Computer Studies, Hlaing Township along master course work and research.

I also like to acknowledge **all my patient teachers** who taught me throughout the master's degree course and **my friends** who explain for their collaboration, discussion lesson and, then wish to express my gratitude to **my beloved parents and**

my brother for their main support and encouragement to fulfill my wish along my life. I also would like to thank all of my school teachers with my job who teach course instead of me from B.E.H.S (8) and B.E.M.S (177) at Pathein Township, for their supports during my academic study.

STATEMENT OF ORIGINALITY

I here by certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

------------------------------ -------------------------

Date Swe Zin Aung

ABSTRACT

Grapheme to Phoneme Conversion (G2P) is the task of automatically generated the pronunciation on a given input word. Pronunciation dictionary is one of the most important things for building automatic speech recognition (ASR) and Text-to-Speech Systems (TTS). The G2P conversion model is implemented for foreign words in Myanmar language using n-gram language modeling and Weighted Finite State Transducer (WFST) based approach. Firstly, Pronunciation Dictionary was built for foreign words in Myanmar language. The alignment of corresponding grapheme to phoneme sequence pair had been generated on the dictionary. A joint n-gram model was trained based on joint grapheme to phoneme chunks aligned during the training process. Finally, the joint n-gram model is converted to an equivalent Weighted Finite State Transducer (WFST). The performance of the model has been evaluated based on Phoneme Error Rate (PER). To ensure the validity of manually prepared pronunciation dictionary and the consistency of the performance of the G2P model, 10-fold cross validation was applied on the data and 2.36% in average Phoneme Error Rate (PER) was obtained for a test set.

TABLE OF CONTENTS

CHAPTER 3 BACKGROUND THEORY

CHAPTER 4 BUILDING PRONUNCIATION DICTIONARY FOR FOREIGN WORDS

CHAPTER 5 BUILDING GRAPHEME TO PHONEME CONVERSION MODEL

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1
INTRODUCTION

Grapheme-to-Phoneme (G2P) conversion is the process of automatically generated phoneme symbols form of unseen words of its pronunciation. For example, given a Foreign word in Myanmar Language "ကဖေး",the process is to generate its pronunciation "**k a- hp ei**:''.It is one of the important things for developing automatic speech recognition (ASR) and text-to-speech (TTS) development for Myanmar language. G2P conversion model is intended to get the correct pronunciation of a given input sequence.

In this thesis, a G2P conversion model is presented and implemented for foreign words in Myanmar language using Joint Sequence N-gram and Weighted Finite State Transducer approach. It supports many-to-many alignment using expectation maximization (EM) algorithm, and sequence modeling using statistical n-gram model. The G2P conversion model for foreign words in Myanmar Language can be divided in two parts. The first part is building pronunciation dictionary and second part is implementing the G2P conversion model.

Firstly, pronunciation dictionary was built for foreign words in Myanmar language. This step is called data preparation. Pronunciation dictionary involves collecting a suitable pronunciation lexicon for training. Pronunciation dictionary was built manually for foreign words in Myanmar Language.

After that G2P conversion model was trained typically in the training process. It can be broken down into three sub-processes (1) Sequence alignment, (2) Model training and, (3) Decoding. This first step is to align training lexicon, segment a pronunciation dictionary and to generate a mapping between the graphemes and phonemes in the lexicon. In this step, auto generated alignment algorithm is applied for generation G2P aligned chunks. In second step, sequence alignment was trained based on N-Gram and Weight Finite State Transducer. In final step, pronunciations for previously unseen words are predicted by using weighted composition to compute the intersection of the WFST representation of the target word and the trained G2P model.

1.1 Motivation of the Thesis

Foreign Words and name entities can cause the more error rate than regular words across all models [1].There is no pronunciation dictionary for foreign words in Myanmar language. So, there is no G2P conversion model for foreign words. It needs to generate correct pronunciation of foreign words in TTS and ASR.

1.2 Objectives of the Thesis

The main objective of the study is to get the correct pronunciation of foreign words in the written Myanmar text. The specific objectives are

(i) To build manual pronunciation dictionary for foreign words in Myanmar language.

(ii) To train the G2P model that is used to generate the pronunciation of new foreign words.

(iii) To apply n-grams and WFST approach in modeling G2P conversion for foreign words.

(iv) To evaluate the performance of WFST based G2P for foreign words in Myanmar language.

1.3 Overview of the Propose System

The Weighted Finite State Transducer based G2P model was developed for automatic G2P conversion of foreign words in Myanmar language that was composed of three processes such as:

(1) G2P alignment for input sequence that to align the grapheme and phoneme chunks pairs in a training dictionary.

(2) Model training by WFST based on N-gram language model that process is to produce a model able to generate new pronunciations for foreign words in Myanmar Language.

(3) Decoding is to find the best pronunciation by the model.

The pronunciation dictionary of foreign words was built and it consists of 34,000 entries. Joint N-gram language modeling and Weighted Finite State Transducer (WFST) based approach are applied in modeling G2P conversion. The performance of G2P conversion model is evaluated in terms of Phoneme Error Rate (PER).10-fold cross validation is done on pronunciation dictionary.

1.4 Related Works

Soky et al. [14] researched Khmer G2P conversion based on weighted finite state transducer (WFST). In this paper, they presented performance of G2P on Khmer language pronunciation dictionary by comparing Rule-based and WFST techniques. The performance of the WFST based G2P is more accurate than rule-based G2P technique. The result were obtained 3.49% in Phoneme Error Rate (PER) or 2.98% in word error rate (WER) for test set.

A. M. Hlaing et al. [1] analyzes sequence to sequence models in G2P conversion for Myanmar language. In this paper Myanmar pronunciation dictionary was built that is applied on sequence to sequence models such as joint sequence model, Transformer, simple encoder-decoder, and enabled encoder-decoder models that were measured by in terms of phoneme error rate(PER) and word error rate(WER). The PER and WER were gained 1.7% and 1.0 % respectively.

Y. K. Thu et al. [17] investigated Myanmar G2P conversion by using four Myanmar syllable pronunciation patterns as features that can be used in a Conditional Random Field (CRF) approach. The results were shown with the Myanmar Language Commission (MLC) test data and the Basic Travel Expression Corpus (BTEC-3) test data. In this paper, all four features gave rise to the highest performance. The word accuracy and phoneme (87.90% and 95.04 %) were achieved by Feature-1234 based on (MLC).

Y. K. Thu et al. [18] examine G2P conversion approaches such as Adaptive Regularization of Weight Vectors (AROW) based structured learning (S-AROW), Conditional Random Field(CRF), Joint-sequence model (JSM), PBSMT, RNN, Support Vector Machine (SVM) based point-wise classification, Weighted Finite State Transducer (WFST) based on manually tagged Myanmar dictionary.G2P conversion models such as CRF, PBSMT and WFST approaches are the best performing methods. The result of phoneme error rate (PER) was obtained approximately 13% for a testing.

Y. K. Thu et al. [19] researched Myanmar G2P conversion system using phrase-based statistical machine translation (PBSMT) approach. PBSMT was achieved the higher performance than CRF and can rapidly predict the right new pronunciations on the unseen compound words. This approach can also deal with the influence of surrounding words on the pronunciation of a word.

1.5 Organization of the Thesis

This study is composed of seven chapters. Chapter 1 presents the introduction and objectives of the study. It also describes the motivation, objectives of this work, overview of the purpose system and related work. Chapter 2 shows the introduction to Myanmar Language. Chapter 3 describes the background theory of Language Model, Finite State Transducer, Weighted Finite Transducer and Smoothing. Chapter 4 describes building pronunciation dictionary for foreign words. Chapter 5 describes how to build Grapheme to Phoneme Conversion and evaluation. Chapter 6 explains the design and implementation of programming modules for Graphical User Interfaces, Evaluation and 10-Fold Cross Validation, the experimental results and Error Analysis are presented. Chapter 7 presents conclusion including challenges, benefit and limitation of the proposed system.

CHAPTER 2

MYANMAR LANGUAGE

This chapter presents the introduction of Myanmar language, the phoneme symbols of consonants, vowels, tones, medial and foreign word in Myanmar language. These symbols are very important to build pronunciation dictionary of conversion foreign word in Myanmar Language.

2.1 Introduction to Myanmar Language

Myanmar language former known as Burmese is the official language of Myanmar and is spoken by 33 million people as a first language and by another 10 million people as a second language. It is a member of the Sino-Tibetan family of languages of which the Tibetan Myanmar subfamily forms a part. Myanmar script derives from Brahmi script [21]. There are basic 12 vowels and 33 consonants and 4 medians in Myanmar language. In Myanmar language, words are formed by combining basic characters with extended characters.

2.2 Phoneme Symbols for Consonants

However pronunciation dictionary needs to extend the dictionary for foreign pronunciations. Myanmar syllables can stand one or more extended characters by combining consonants to form compound words. The first 25 consonants are divided into five classes. They are velars, palatals, alveolars, dentals and labials.

Table 2.1 Phoneme Symbols for Consonants

Group Name	Unaspirated (သိထိလ)	Aspirated (ဓနိတ)	Voiced (လဟု)		Nasal (နိဂ္ဂဟိတ)
Velars	က /k/	ခ /kh/	ဂ /g/	ဃ /g/	င /ng/
Palatals	စ /s/	ဆ /hs/	ဇ /z/	ဈ /z/	ဉ/ည / nj/
Alveolars	ဋ /t/	ဌ /ht/	ဍ /d/	ဎ /d/	ဏ /n/
Dentals	တ /t/	ထ /ht/	ဒ /d/	ဓ /d/	န /n/
Labials	ပ /p/	ဖ /hp/	ဗ /b/	ဘ /b/	မ /m/
Without Group	ယ /j/	ရ /j/r/	လ /l/	ဝ /w/	သ /th/
		ဟ /h/	ဠ /l/	အ /a/	

Table 2.1 shows the groups of characters according to their pronunciation phoneme symbol 33 consonants and the groups are unaspirated, aspirated, voiced and nasal. Many Myanmar syllables containing unaspirated and aspirated consonants are pronounced as voiced consonants depending on the neighboring context.

2.3 Phoneme Symbols for Vowels

Basically, There are 12 vowels in Myanmar writing. These vowels are အ(a) ,အာ(a), အိ(i.) , အီ(i), အု (u.), အူ(u) , အေ(ei) အဲ(e:), အော (o:), အံ့ (an.), အို (ou). The following Table 2.2 shows phoneme symbols for 12 vowels are basically 12 vowels for Myanmar language.

Table 2.2 Phoneme Symbols for Vowels

non-nasalized vowels								nasalized vowels					
အိ	i.	အီ	i	အီး	i:	အစ်	i’	အင်	in	အင့်	in.	အင်း	in:
အေ	ei	အေ့	ei.	အေး	ei:	အိတ်	ei’	အိန်	ein	အိန့်	ein.	အိန်း	ein:
အယ်	e	အယ့်	e.	အဲ	e:	အိုက်	ai’	အိုင်	ain	အိုင့်	ain.	အိုင်း	ain:
အာ	a	အာ့	a.	အား	a:	အတ်	a’	အန်	an	အန့်	an.	အန်း	an:
အော်	o	အော့	o.	အော	o:	အောက်	au’	အောင်	aun	အောင့်	aun.	အောင်း	aun:
အူ	u	အု	u.	အူး	u:	အွတ်	u’	အွန်	un	အွန့်	un.	အွန်း	un:
အို	ou	အို့.	ou.	အိုး	ou:	အုပ်	ou’	အုန်	oun	အုန့်	oun.	အုန်း	oun:

2.4 Tone in Myanmar Language

Myanmar is a tonal language with Creaky, Low and High in Table 2.4. Tones are denoted with different sound or specialized symbols in writing Myanmar scripts text. Vowel quadrilateral of vowels in Myanmar language are described in Table 2.3 [2].

Table 2.3 Vowel Quadrilateral for Myanmar Vowels

	Front	Central	Back
High	အီ /i/		အူ /u/
Mild high	အေ/e/		အို /o/
Mild low	အယ် /ɛ/		အော /ɔ/
Low		အာ/a/	

Table 2.4 Tone in Myanmar Language

Tone	MLC	Phonation	Length	Myanmar	Meaning
Tone 1	k a	Normal	moderate	ကာ	cover
Tone 2	k a:	Breathy	long	ကား	car
Tone 3	k a.	Creaky	short	က	dance
Tone 4	k a’	Final Glottal stop	abrupt	ကဝ်	stick

2.5. Phoneme symbols for Medial

There are four basic phoneme symbols for medial in Myanmar Language (-ျ/ -ြ / -ွ / -ှ). These Myanmar script can be joined with appropriate symbols of the 33 consonants. Myanmar script permits the palatal medial.

Table 2.5 Phoneme Symbols for Medial

base	Phoneme Symbols	Example	
ျ	j	မျ	m j a.
ြ	j	မြ	m j a.
ွ	w	မွ	m w a.
ှ	h	မှ	h m a.

2.6 Phoneme Symbols for Foreign Words

Some foreign pronunciations have to be expressed by representing foreign words because Myanmar pronunciations dictionary do not contain phoneme symbols. So, we also use phoneme symbols for foreign words, in addition to the phoneme symbols defined in MLC. The following table shows about phoneme symbols for foreign word. For example, the Myanmar phonetic representation of the foreign name “နယူးစ်” “News” is “n a- j u: S” and “ဆိုလ်း” “Seoul” is “hs ou: L”.

Table 2.6 Phoneme Symbols for Foreign Words

Foreign	Phoneme Symbols	Foreign	Phoneme Symbols
(က်)	**K**	(ခ်)	**KH**
(ဂ်)	**G**	(စ်)	**S**
(ဆ်)	**HS**	(ဇ်)	**Z**
(တ်)	**T**	(ထ်)	**HT**
(ဒ်)	**D**	(ဓ်)	**D**
(န်)	**N**	(ပ်)	**P**
(ဖ်)	**HP**	(ဗ်)	**B**
(ဘ်)	**B**	(မ်)	**M**
(ယ်)	**Y**	(ရ်)	**R**
(လ်)	**L**	(ဝ်)	**W**
(သ်)	**TH**	(ဟ်)	**H**
(ချ်)	**CH**	(ရှ်)	**SH**

CHAPTER 3
BACKGROUND THEORY

This chapter describes N-gram language model, smoothing and types of smoothing, Kneser-Ney Smoothing that is used to solve zero probabilities problem with MITLM model. The details of Weight Finite Transducer and Finite State transducer are described how to decode using Weight Finite Transducer to produce phoneme symbols in Myanmar Language.

3.1 N-gram Language Model

An N-gram [4] is the pair of a word sequence containing number of words and its according count, based on the occurrences of a given sequence in a corpus. N-gram model predicts the most possible word that might follow this sequence in a corpus. It is a probabilistic model that was trained on a corpus of input unseen text. This model is useful in NLP applications including speech recognition, machine translation, and predictive text input, dictionary. An N-gram language model is built by counting word sequences occur in corpus text and then calculate the probabilities. Since a simple N-gram model has limitations, improvements are often made via smoothing. Joint-sequence model [11] is the most popular among many different approaches that have been proposed.

An N-gram model is one type of a Language Model (LM) that is finding the probability over word sequences. A model relies on a word occurs and cannot consider at previous words is called unigram**.** If a model considers only from the previous word to predict the current word, then it is called bigram and also if two previous words are considered situation then it is a trigram model.

Bigram grams and N gram formula is presented below:

$$\textbf{Bigram:} \quad P(w_n \mid w_{n-1}) = \frac{C(w_{n-1}w_n)}{C(w_{n-1})}$$

$$\textbf{N-gram:} \quad P(w_n \mid w_{n-N+1}^{n-1}) = \frac{C(w_{n-N+1}^{n-1}w_n)}{C(w_{n-N+1}^{n-1})} \qquad \text{……….(3.1)}$$

where,

W_n is current state

W_{n-1} refers to the previous words from current state

N describes the number of words in the corpus

C is the count of times the word sequence occurs in the corpus

3.2 Smoothing

Smoothing [5] enhances the accuracy in statistical models when applied to data with sparse the distribution by redistributing weight from high probability regions to zero probability regions. The process of smoothing is very important in natural language processing that case zero or near to zero probabilities means that word do not exist in the vocabulary. This problem is called overfitting. Smoothing methods can overcome overfitting problem. It was quite possible that some word sequences were occurred in test data that could never see during training process. When this happens, the probability of the sequence equals zero. Evaluation is also difficult because perplexity metric becomes infinite. The use of smoothing process is to handle data spare words that are absent in the training set and out of vocabulary words and to improve the accuracy of model. There are four types of smoothing. They are

1. Laplace

   Add one only to the count of the words.

2. Additive

   Adding delta, “δ”. Instead of adding 1 as like in Laplace smoothing, a delta (δ) value is added.

3. Interpolation

   It interpolates lower level n-grams with higher level n-grams such as interpolating unigram probabilities with bigram probabilities to prioritize some bigrams that should have higher counts than the assumed alpha value or 1 in the prior two methods.

4. Kneser-Ney

   It is observed that the count of n-grams is discounted by a constant/absolute value such as 0.6.The absolute discounting is applied to the count of n-grams in addition to adding the product of interpolation weight and probability of word to appear as novel continuation.

3.3 Kneser-Ney Smoothing

Kneser-Ney Smoothing [10, 12] is a good baseline method and it is based on absolute Kneser and Ney that is intended to improve backoff weight and discounting. It is based on the concept of absolute discounting that is removed from all non-zero counts. It improves on absolute discounting by estimating the count of a word in a new context based on the number of different contexts. It makes use of both higher and lower-order language models, calculate some probability value from 4-grams or 3-grams to simpler unigram models.

The formula for absolute-discounting smoothing as applied to a bigram language model is presented below:

$$P_{KN}(w_i \mid w_{i-1}) = \begin{cases} \dfrac{C(w_{i-1}\, w_i) - D}{C(w_{i-1})}, & \text{if } C(w_{i-1}\, w_i) > 0 \\ \propto (w_i) \dfrac{|\{w_{i-1} : C(w_{i-1}\, w_i) > 0\}|}{\sum_{w_i} |\{w_{i-1} : C(w_{i-1}\, w_i) > 0\}|} & \text{otherwise} \end{cases} \quad \text{............. (3.2)}$$

where ,

W_i is current state

W_{i-1} refers to the previous words from current state

W describes the number of words in the corpus

C is the count of times the word sequence occurs in the corpus

D=Discount values

P_{KN}=Probabilities of Kneser-Ney

3.4 Finite State Automata

A finite state machine sometimes called a finite state automaton [13] is a computation model to recognize pattern. It is used by programmers, mathematicians, engineers and other professionals to describe a mathematical model that can be implemented with hardware or software and can be used to simulate sequential logic and some computer programs for the system.

The finite state automata or finite state machine has five elements or tuples that has a set of states and rules moving from one state to another state base on the applied input symbol. The following describes are five features of automata. They are

1. Input
2. Output
3. States of automata

4. State relation
5. Output relation

There are two types of finite state machines that are deterministic finite state machines and non-deterministic finite state machines, sometime often called deterministic finite automata and non-deterministic finite automata.

3.4.1 Deterministic Finite Automata (DFA)

In DFA, for each input symbol, one can determine the state to which the machine will move. Hence, it is called Deterministic Automaton. As it has a finite number of states, the machine is called Deterministic Finite Machine or Deterministic Finite Automaton**.**

Deterministic Finite Automata composes of five tuples that are {Q, Σ, q, F, δ}.

Q: is defined set of all states are vertices.

Σ: set of input symbols that machine takes as input alphabet.

q: Initial state is starting state of a machine with an empty symbol single incoming arc.

F: is defined set of final state with double circles.

δ: Transition Function is defined as δ that are arcs labeled.

3.4.2 Nondeterministic Finite Automata (NFA)

NFA does not have restrictions and the transition function is expressed by the empty string ε. An ε-transition is a transition without input symbol and is represented with by an arrow labeled "ε" in state diagrams. ε-transitions provide current states are not precisely known modeling systems that is not clear whether the current state after processing some ,then an ε-transition was added between these two states for input string should be *q* or *q'*, in both simultaneously states.

3.1 Weighted Finite State Transducer

The WFST [14] is a popularity part in NLP communities and the speech in recent years because it provides decoding techniques, wide range of various model and also WFST is suitable tools for the process of combination, optimization, training and decoding .WFST is the similar unweighted FSA. Finite State Transducer [6] is a finite automaton whose state transitions are regarded as labeled with both input and output symbols. Transition in WFST machine is intended to encode a weight, and both the input labels in the FSA and output label. When the weight is completed valid path, it through the WFST is computing the product of the arc weights along the path and, the output labels regarded a second output through the machine.

Weighted transducers are represented the probabilistic finite-state models in speech processing. Weighted transducers are used such as text, speech and image processing and to assign different pronunciations to the same word. Transducers can be used to define a mapping between word and phoneme sequence of information sources. WFST and FST are automata that input label is augmented as transition and generate possible new alphabet as output label and carries weight of their element.

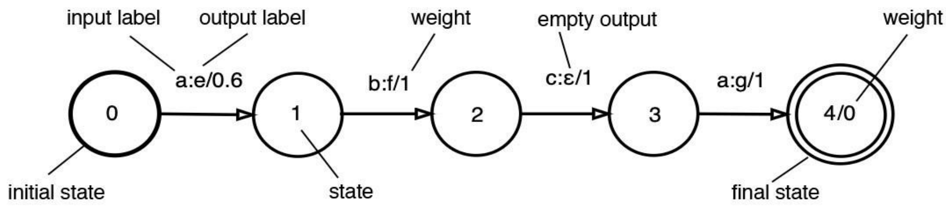

Figure 3.1 Weighted Finite State Transducer

In Figure 3.1, WFST is composed of three state .They are initial state, state and final state. The start state is called initial state and the end state with double circle is called final state. The state between the initial state and the final state is called state. States are “1”, “2”, “3”. Initial state is “0” and final state is “4”. Input label are the input grapheme and output label are the output phoneme. Input label are “a”, “b”, “c” and output label are “e”, “f”, “g”. If the output phoneme is empty, it is called empty output .The symbol “/” followed by the value is called weight such as “0.6”, “1” and “0”.

CHAPTER 4

BUILDING PRONUNCIATION DICTIONARY FOR FOREIGN WORDS

This chapter describes how to build pronunciation dictionary using phoneme symbols that describes in Chapter 2.The preparation of step by step process describes how to prepare data for training. This step involves manual check and repair and foreign words pronunciation dictionary.

4.1 Building Pronunciation Dictionary

For implementing Weighted Finite State Transducer model for G2P conversion, at least, a pronunciation dictionary is required. Therefore, a foreign pronunciation dictionary was built for modeling G2P. The size of its vocabulary is 34,000. As much as possible pronunciations of each word are collected in the dictionary. This dictionary can also be applied in Myanmar TTS and ASR application.

Figure 4.1 describes about the process flow diagram of building pronunciation dictionary. The following steps were applied for building pronunciation dictionary for the statistical G2P conversion model based on training.

(1) Firstly, pronunciation dictionary was built and the transliterated Myanmar Words are extracted from English-Myanmar Transliteration [20].

(2) In second step, the transliterated foreign words were tagged by using Sequitur G2P model to generate pronunciation dictionary.

(3) Finally, phonemes are manually checked and repaired to build final pronunciation dictionary. This step is tough and time consuming. Final pronunciation must be trained into the G2P conversion model.

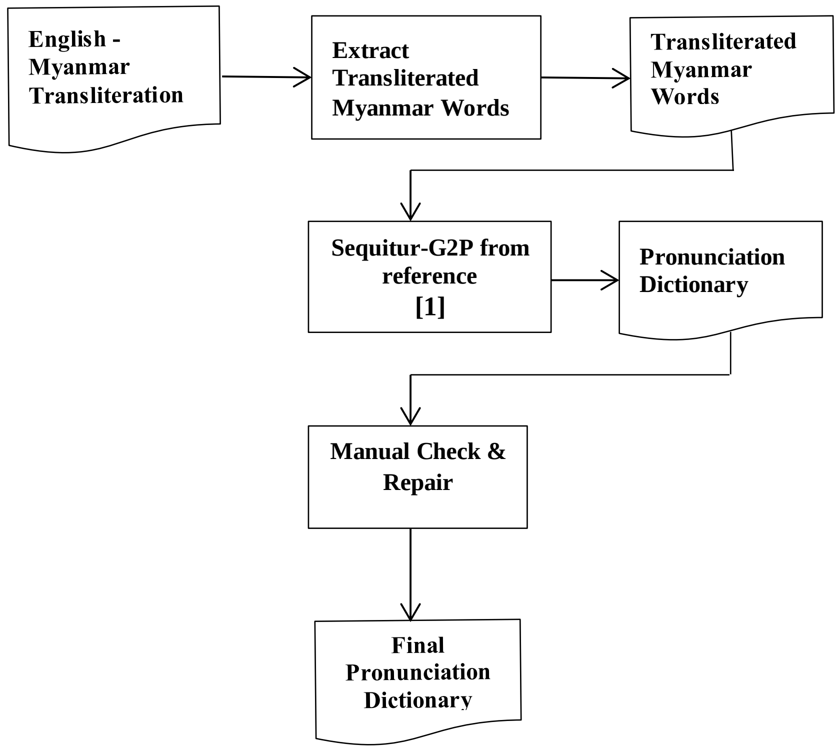

Figure 4.1 Process Flow of Building Pronunciation Dictionary

4.2 Manual Check and Repair

The next step is to check up and repair data. The input data is automated tagged by Sequitur G2P model .The output data is not perfect generated. So, the generated phoneme symbols have to be checked and repaired for training. Before starting the training and testing, some phoneme symbols are added, small letter symbols changed to capital symbols, and some are deleted and inserted as shown in the Table 4.1.

Table 4.1 Example of Manual Check and Repair

Foreign Words	**Sequitur G2P**	**Manual Check & Repair**
ဆော့ဖ်ဝဲ	hs o. **P** w e:	hs o. **HP** w e:
ကာဗွန်ဒိုင်အောက်ဆိုဒ်	k a b un d ain au' hs ai’	k a b un d ain au' hs ai’ **D**
စီအက်ဖ်အို	s i e' H P ou	s i e' HP ou
ဆူဖ်ရမ်	hs u hp r an	hs u HP r an

4.3 Pronunciation Dictionary for Foreign Words

To examine the model of the G2P conversion system, building pronunciation dictionary is an important work to explore phoneme symbol for foreign words in Myanmar based on n gram and weight finite state techniques. There was published standard Myanmar pronunciation dictionary, Myanmar Language Commission (MLC) dictionary [7], which has about 27,300 words .The phoneme symbols was used in the system. Therefore, pronunciation dictionary of foreign words in Myanmar language was built to provide TTS system and ASR development. The size of pronunciation dictionary is 34,000 entries that contain possible pronunciation of foreign words are collected in the dictionary.

Commonly used foreign words are mainly focused in collecting and preparing the data. The lexicon in pronunciation dictionary had been built common loan words and common foreign name entities. The collected data were used for training G2P conversion model apply N-gram and WFST based G2P approaches. Finally, the large foreign pronunciation dictionary including the 34,000 pairs of words and pronunciations was achieved. Examples of selected words and their pronunciations in the pronunciation dictionary are described in Table 4.2. To the best of the knowledge, this is the first foreign pronunciation dictionary for Myanmar language.

Table 4.2 Example Entry of Foreign Words Pronunciation Dictionary

Foreign Words	In English	Pronunciation for Foreign Words
ဒီဇိုင်နာ	Designer	d a- z ain n a
မိုက်ကရိုဖုန်း	Micro Phone	m ai’ kh a- r o hp oun:
ခရစ်ယာန်	Christian	kh a- r i’ j an
နျူးကလီးယား	Nuclear	n a- j u: k a– l i: j a:
ရုရှား	Russia	r a- sh a:
ဝဘ်ဆိုက်	Website	w e’ hs ai’
ဆော့လ်ဝဲ	Software	hs o. L w e:
စင်တာ	Center	s in t a
ကဖေး	Cafe	k a- hp ei:
တစ်သျှူး	Tissue	t i’ sh u:
ဆမ်းဒဝစ်ဒ်	Sandwich	hs an: d a- w i’

Foreign Words	In English	Pronunciation for Foreign Words
ဆန်ဖရန် စစ္စကို	San Francisco	hs an hp a- r an s i’ s a- k o
ဒီဇိုင်း	Design	d a- z ain:
ဖေ့စ်ဘုတ်	Facebook	hp ei. S b ou’
အော်စကာ	Oscar	o s a- k a
အီးမေးလ်	Email	i: m ei: L
ဒေါင်းလုဒ်	Download	d aun: l ou’
ပေရာယ်	April	ei p a- r e
ဩစတြေးယား	Australia	o s a- t ei: j a:
လက်တော့ပ်	Laptop	l e’ t o. P
ပရိုဂျူဆာ	Producer	p a- r ou gy u hs a

In Figure 4.3, there are a total number of 25 unique phoneme symbols for consonant. There are a total number of 50 unique phoneme symbols for vowels and 19 unique phoneme symbols for foreign. Model was used a total number of 94 unique grapheme-phoneme pairs to cover pronunciation dictionary.

Table 4.3 Unique Phoneme Symbols in Pronunciation Dictionary

Phoneme Symbols	Number of Unique Phoneme Symbols
Consonant	25
Vowel	50
Foreign	19
Total	**94**

In Figure 4.2, the statistics of 133,581 phoneme symbols were shown for consonant to show the coverage of consonants in pronunciation dictionary. Phoneme symbols that “b”, “m”, “j”, “r”, “k” and “l” was mostly found in G2P conversion machine. Total numbers of phoneme symbols for consonants are 11746 numbers that are the highest amount among all symbols and the phoneme symbol b consonants are the second largest numbers that have 11241consonats.Some symbols such as “ky”, “kh”, “nj”, “ht” and “hp” were less used in training data set.

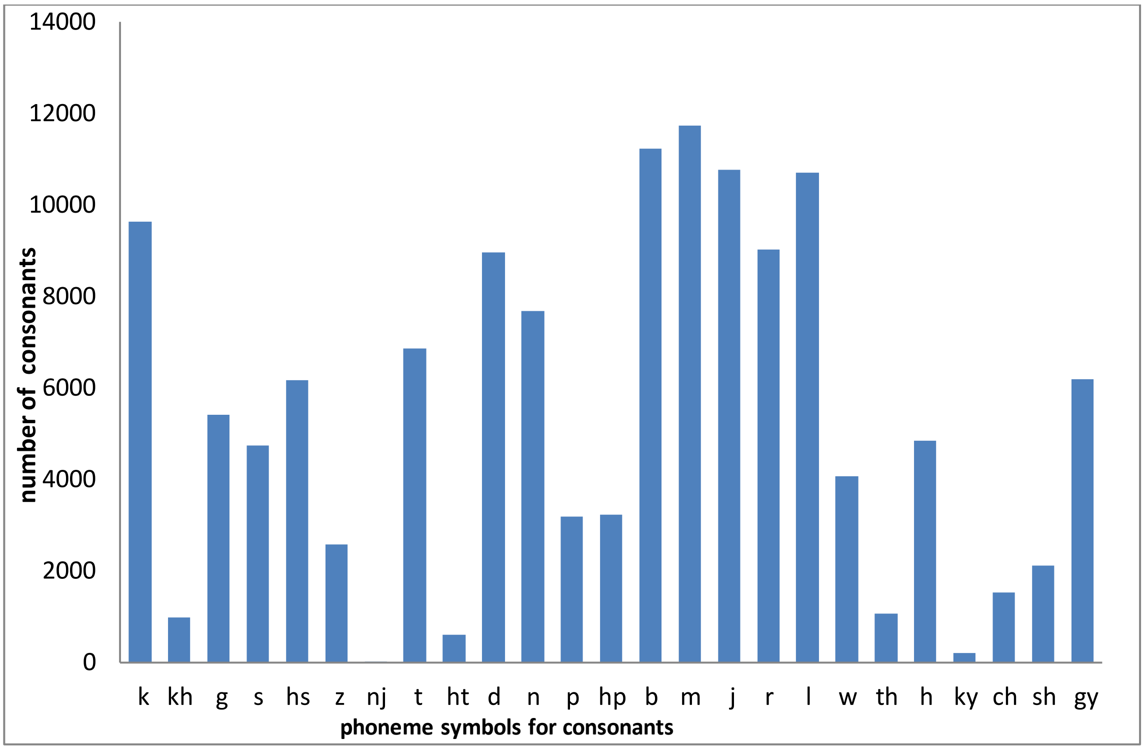

Figure 4.2 Statistics of Phoneme Symbols for Consonants

The next chart in Figure 4.3 describes the count of phoneme symbols for vowels. The largest numbers of vowel that phoneme symbol "i" is 21450 and these symbols are mostly used in foreign dictionary. Phoneme symbol, "a-" 14775 numbers of vowels was found in Figure 4.4. This symbol can occur sometimes when grapheme symbol was read, these tone of grapheme appeared "central" tone denoted with the phoneme symbol "a-" .This place can mostly cause vowel errors for pronunciation dictionary.

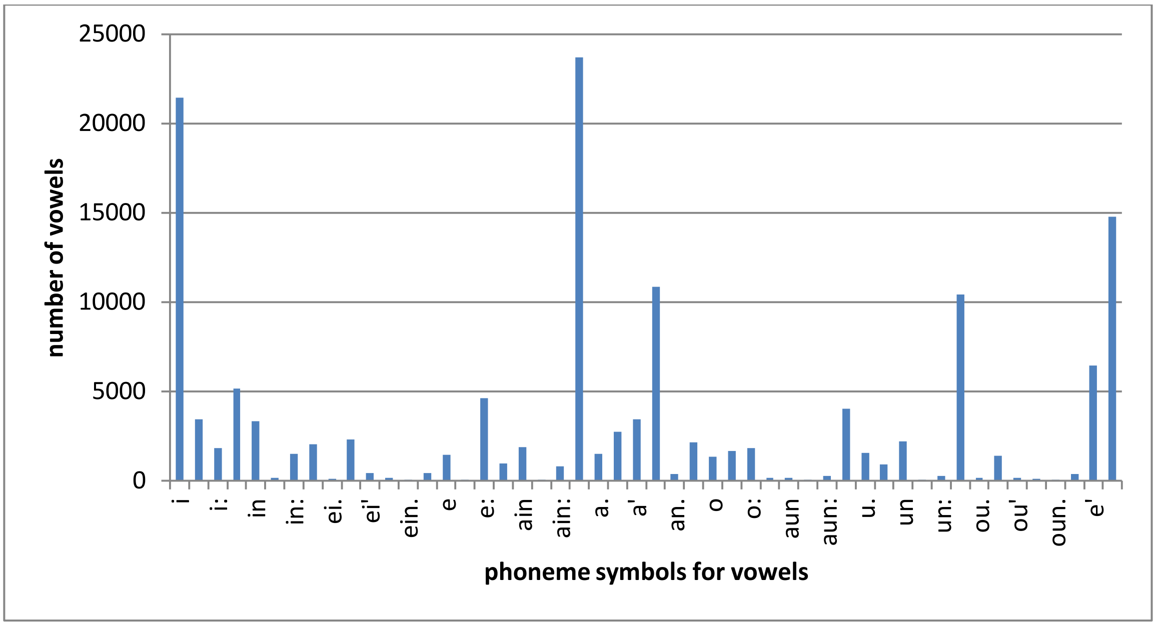

Figure 4.3 Statistics of Phoneme Symbols for Vowels

Figure 4.4 describes the number of statistics phoneme symbols for foreign symbols that have 13823 phoneme symbols. The statistics of foreign symbols for phoneme symbols are less used in the system. Foreign symbol of "L" for phoneme symbols was mostly occur in dictionary. The foreign symbol "L" had 5908 numbers of foreign. This symbol can occur such as the name "John", "ဂျွန်(လ်)" that "L" foreign symbol is denoted for "လ်". The G2P machine will be generated the pronunciation "gy un L" for this grapheme word "ဂျွန်(လ်)" pronunciation.

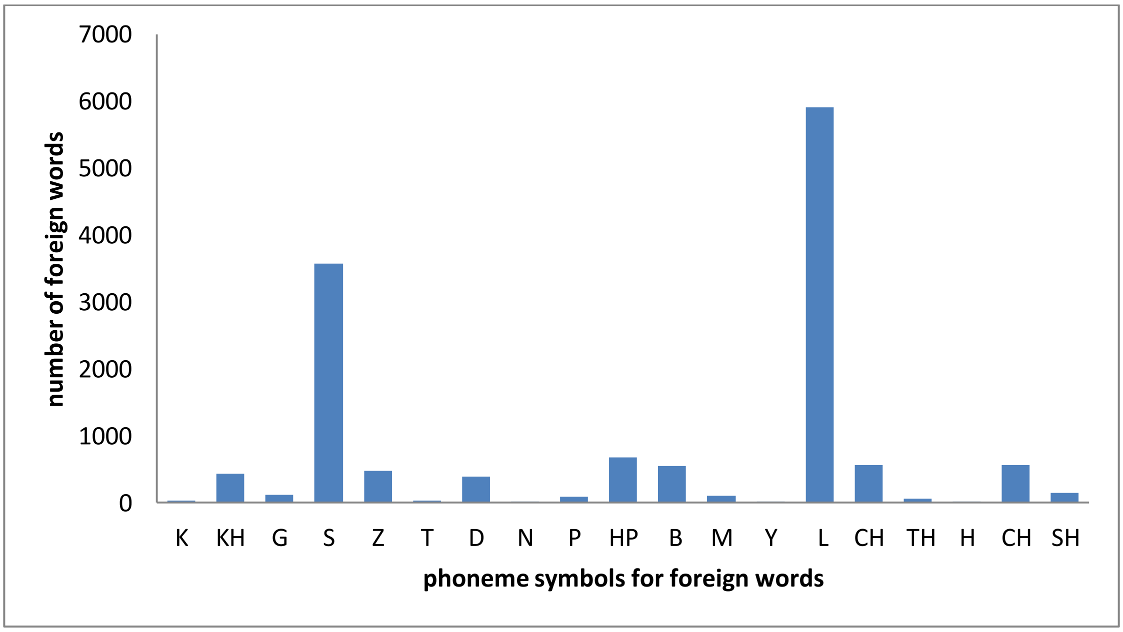

Figure 4.4 Statistics of Phoneme Symbols for Foreign Words

CHAPTER 5

BUILDING GRAPHEME TO PHONEME CONVERSION MODEL

This chapter presents about building grapheme to phoneme conversion model and how to align for grapheme to phoneme alignment. It also describes the detail process of N-gram language Model with MITLM trained on G2P alignment pairs and to calculate of 2 grams probabilities and format ARPA file that generated n-gram language model and covert ARPA format into WFST to construct automata with grapheme of their phoneme symbols.

5.1 Building Grapheme to Phoneme Conversion Model

Foreign Words in Myanmar Language G2P conversion approach utilized in this work and implemented in Phonetisaurus [9] using Open FST [3] frameworks. The training procedure is then,

1. Convert aligned sequence pairs to sequences of aligned joint label pairs, $(g1{:}p1, g2{:}p2, ..., g_n{:}p)$;
2. Train an Joint Sequence N-gram model by generated G⇔P alignment file
3. Convert the Joint Sequence N-gram model into a WFST.

The final pronunciation dictionary that contains 34,000 tagged data had been aligned by using automatic generated alignment algorithm in training process. The model automatically generated G⇔P alignment file. These alignment output file was trained by using n-gram language model to produce ARPA output file. The ARPA output file had been converted to an equivalent Weighted Finite State Transducer model. Finally, model was obtained that WFST based on G2P conversion model to get the pronunciation of foreign word in Myanmar language. After the system had been generated model, the model will be tested. So, the user input unseen data foreign word to the system. If the user wants to test the foreign word “ကဖေး” in Myanmar language, the model will generate it’s pronunciation with phoneme symbols of this “k a- hp ei:” in testing process.

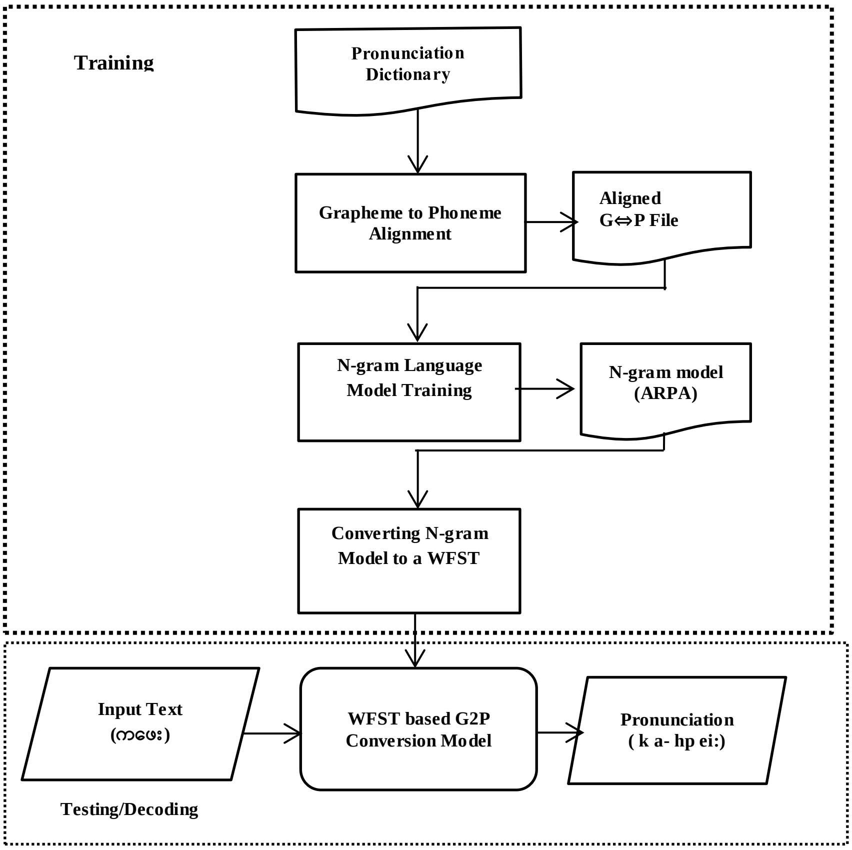

Figure 5.1 System Flow for Proposed System

5.2 Grapheme-to-Phoneme Alignment

G2P system in first training process involves aligning the corresponding input grapheme and phoneme sequences in training pronunciation dictionary. This process is based on the EM driven multiple-to-multiple alignment algorithm described in [15] and proposed [16].The alignment algorithm utilized includes two process :(1) G2P alignment is imposed such that one-to-one relationships, only multiple-to-one and one-to-multiple relationships are considered during training. (2) a joint sequence alignment chuck is constructed for each input entry during initialization. In first training process, G2P system involves aligning the corresponding input foreign grapheme and phoneme symbols in training pronunciation dictionary.G2P alignment is capable of learning align G↔P relationships like က →/k. The result of which is a chunk of aligned, joint sequences. As an example, four foreign words are used with their pronunciations shown in Table 5.1 and Table 5.2 shows the results of aligned chunks of Grapheme ↔Phoneme.

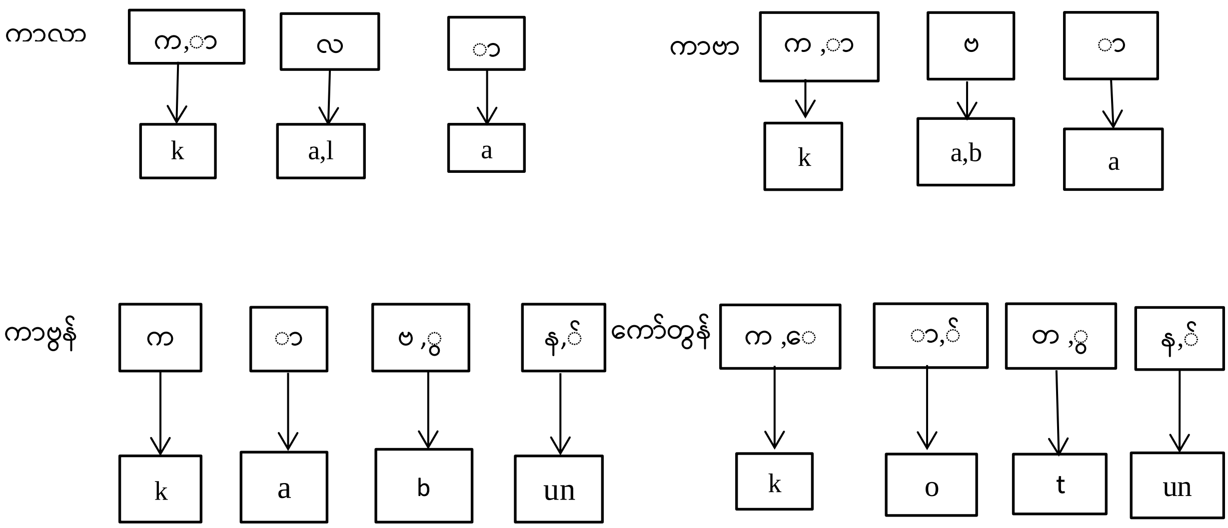

Figure 5.2 Example of Alignment

G2P alignment is imposed such as many-to-one and one –to-many arcs will be considered during training. In Figure 5.2, A ‘,’ indicates a one-to-many relationship. For example, the foreign word “ကာလာ” grapheme is divided into three chunks. The first chunk, က,ာ →*/k* called many –to-one, the second chunk လ →*/a, /l* called one-many and last chunk is called ာ →*/ a* called one –to-one.

Table 5.1 Entry of Foreign Pronunciation Dictionary

Foreign Words	Pronunciation
ကာလာ	k a l a
ကာဗာ	k a b a
ကာဗွန်	k a b un
ကော်တွန်	k o t un

Table 5.2 Aligned chunks of G↔P

Aligned chunks of G↔P
က,ာ||| k လ||| a,l ာ||| a
က,ာ||| k ဗ||| a,b ာ|||a
က||| k ာ||| a ဗ,ွ||| b န,်|||un
က,ေ|||k ာ,်|||o တ,ွ|||t န,်|||un

5.3 N-gram Language Model with MITLM

The MIT language modeling toolkit (MITLM)[8] toolkit is used to build n-gram language model based on training data modified with Kneser-Ney smoothing. After once the input pronunciation dictionary has been successfully aligned, the next step is to train a model that can be used to produce pronunciation hypotheses for previously unseen words. The training approach in the proposed system is identified that the ‘words’ are joint G↔P chunks learned during the alignment process. This means that any standard statistical language modeling toolkit may be used to train a joint n-gram model for the proposed system. Language Model was trained on pronunciation dictionary described in Chapter 4 to produce phoneme symbol. Table 5.3 shows count of words in the corpus for 4 words such as “ကာလာ”, “ကာဗာ”,ကာဗွန်”,”ကော်တွန်”.

Table 5.3 Count of words in the Corpus for 4 Words

က|ာ}k=2	လ}a|l=1	ာ}a=3
က}k=1	ဗ}a|b=1	န|်}un=2
က|ေ}k=1	ဗ|ွ}b=1	ာ|်}o=1
တ|ွ}t=1		

Calculation 2-Grams Probability

$$P\,(\text{လ||| a | က|ာ||| k}) = \frac{C(\text{က|ာ|||a}) - D}{C(\text{က|ာ|||k})} = \frac{1-0.6}{2} = \text{Log } 0.2 = -\,0.6987000$$

$$P\,(\text{ဗ||| a|b | က|ာ||| k}) = \frac{C(\text{က|ာ||| k ဗ|||a|b}) - D}{C(\text{က|ာ|||}k)} = \frac{1-0.6}{2} = \text{Log } 0.2 = -\,0.6987000$$

$$P\,(\text{ာ|်|||o | က|ေ ||| k}) = \frac{C(\text{က|ေ||| k ာ|်||| o}) - D}{C(\text{က|ေ|||}K)} = \frac{1-0.6}{1} = \text{Log } 0.4 = -\,0.3979400$$

$$P\,(\text{ာ|||a | က||| k}) = \frac{C(\text{က|||k | ာ|||a}) - D}{C(\text{က|||}k))} = \frac{1-0.6}{1} = \text{Log } 0.4 = -\,0.3979400$$

$$P\,(\text{န|်|||un | တ|ွ||| t}) = \frac{C(\text{တ|ွ}\}\text{t န|်|||un}) - D}{C(\text{တ|ွ|||t})} = \frac{1-0.6}{1} = \text{Log } 0.4 = -\,0397400$$

P (နှ|င်|||un | ဗ|ွ||| b) = $\frac{C(\text{ဗ|ွ|||b နှ|င်||| u n})-D}{C(\text{ဗ|ွ|||b})}$ = $\frac{1-0.6}{1}$ = Log 0.4= - 0.3979400

P (ာ|||a | ဗ||| a|b) = $\frac{C(\text{ဗ|||a |b ာ|||a})-D}{C(\text{ဗ|||a |b })}$ = $\frac{1-0.6}{1}$ = Log 0.4= - 0397400

P (ာ |||a | လ||| a|l) = $\frac{C(\text{လ|||a|l })-D}{C(\text{လ|||a|l })}$ = $\frac{1-0.6}{1}$ = Log 0.4= - 0397400

P (တ|ွ ||| t | ာ|င်||| o) = $\frac{C(\text{ာ|င်|||o တ|ွ}t})-D}{C(\text{ာ|င်|||o})}$ = $\frac{1-0.6}{1}$ = Log 0.4=- 0.3979400

P (ဗ|ွ|||b | ာ||| a) = $\frac{C(\text{ာ|||a ဗ|ွ|||b})-D}{C(\text{ာ|||a})}$ = $\frac{1-0.6}{3}$ = Log 0.133= -0.875072

data\ngram 2=15

\2-grams:

-1.150537	<s> က|ာ}k	2.649686711
-0.812365	<s> က|ေ}k	1.870876595
-0.812365	<s> က}k	1.870876595
-0.651243	က|ာ}k ဗ}a|b	1.499812629
-0.651243	က|ာ}k လ}a|l	1.499812629
-0.409048	က|ေ}k ာ|င်}o	0.942037544
-0.425871	က}k ာ}a	0.980780913
-0.377238	တ|ွ}t နှ|င်}un	0.868779114
-0.888425	နှ|င်}un </s>	2.046042775
-0.377238	ဗ|ွ}b နှ|င်}un	0.868779114
-0.425871	ဗ}a|b ာ}a	0.980780913
-0.425871	လ}a|l ာ}a	0.980780913
-0.409048	ာ|င်}o တ|ွ}t	0.942037544
-0.939578	ာ}a </s>	2.163848134
-0.729247	ာ}a ဗ|ွ}b	1.679455841

\end\

Figure 5.3 ARPA Format with 2 Grams Language Model

The ARPA file is the output of n-gram language model file that predicts the backoff weight value based on n-gram from the total counts. And, ARPA file contains unigram, bigram and trigrams of words. In Figure 5.3, the left of values with (-) sign are log base 10 probability based on n-gram, for example (-1.150537).

The right positive values are converted value from log base 10 into log base e, for example (2.649686711) .That will be explained the next sub title.

```
\data\
ngram 3=14
\3-grams:
    -0.651243    <s> က|ာ}k ဗ}a|b          1.499812629
    -0.651243    <s> က|ာ}k လ}a|l          1.499812629
    -0.409048    <s> က|ေ}k ာ|င်္}o          0.942037544
    -0.425871    <s> က}k ာ}a              0.980780913
    -0.425871    က|ာ}k ဗ}a|b ာ}a          0.980780913
    -0.425871    က|ာ}k လ}a|l ာ}a          0.980780913
    -0.409048    က|ေ}k ာ|င်္}o တ|ွ}t        0.942037544
    -0.729247    က}k ာ}a ဗ|ွ}b            1.679455841
    -0.888425    တ|ွ}t န|င်္}un </s>        2.046042775
    -0.888425    ဗ|ွ}b န|င်္}un </s>        2.046042775
    -0.939578    ဗ}a|b ာ}a </s>           2.163848134
    -0.939578    လ}a|l ာ}a </s>           2.163848134
    -0.377238    ာ|င်္}o တ|ု}t န|င်္}un       0.868779114
```

Figure 5.4 ARPA Format with 3 Grams Language Model

5.4 Convert ARPA to Weighted Finite State Transducer

In the WFST machine, the first step is to generate an alignment lattice for input and output of word-pronunciation pair based on the user input parameters. The final preparatory step is converted the resulting ARPA file contains n-gram model to an equivalent WFST [3].The WFST-based model was intended to produce pronunciation hypotheses for chunk words by first transforming the target word into an equivalent finite-state machine and composing it with the model.

A finite-state transducer (FST) is composed of both the input and output labels with the arc label. A transducer maps with a path from an input tagged word to an output label sequence is called FST. WFST is the weighted version of the finite-state transducer. If an arc does not have nothing input or output, it will be marked with <eps>, ε or "-".It is description in terms of FST components. According to an example in Figure 5.5,

Q = {0,1,2,3,4,5,6,7,8,9,10}
0 is the start state
F = {7, 9}
Where Q= State, F=Final State,
Σ =input graphemes, Δ=output phonemes.
Σ contains က|ာ , လ , ာ , က|ာ , ဗ , ာ , က,ာ,ဗ|ူ , နှ|င် , က|ေ ,ာ|င်, တ|ူ, နှ|င်.
Δ contains at least k, a|l ,a , k, a|b , a , k, a , b,un , k, o,t ,un.

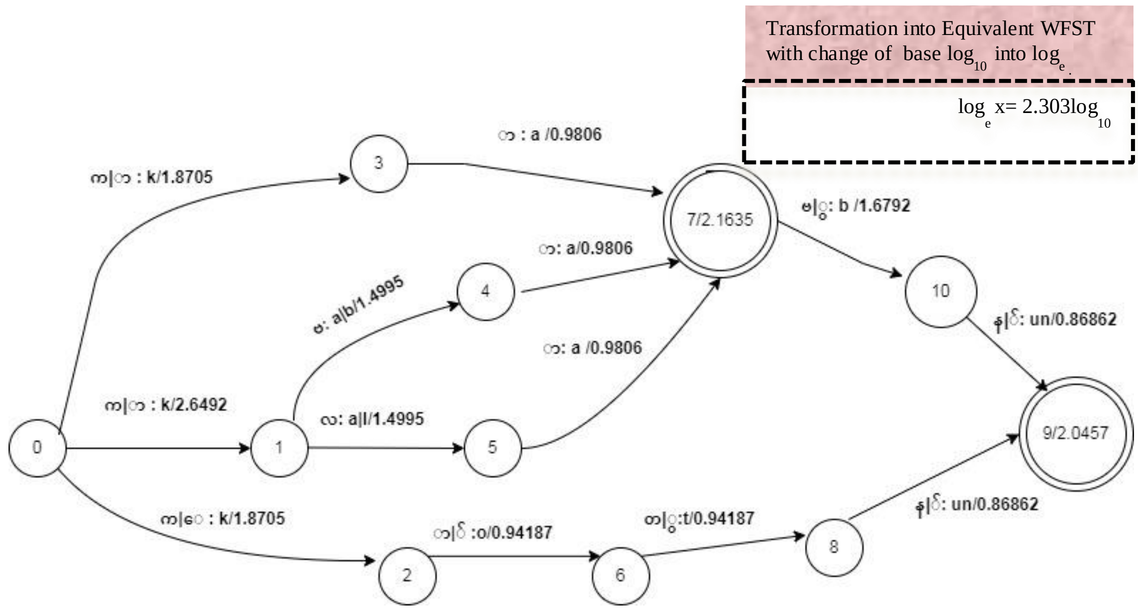

Figure 5.5 Example of Weighted Finite State Transducer For 4 Words

This equation is transformation into equivalent WFST with change of base $\log_{10}$ into $\log_e$

$$\log_e x = 2.303\log_{10} \qquad \text{........................(5.1)}$$

If convert ARPA with n-gram probability that has $\log_{10}$ into WFST with $\log_e$, the log 10 value must be needed to product the value with 2.303 according to the formula $\log_e x= 2.303\log_{10}$. Equation 5.1was used the result of probability based 10 with n-gram convert into Weighted Finite State Transducer with base into $\log_e$

For example, Calculate 2 grams for P (လ}a | က|ာ||| k)= $\frac{\text{C(က|ာ|||k လ}a)} - D}{\text{C(က|ာ}k)}}$

$$= \frac{1-0.6}{2}$$
$$= \text{Log } 0.2$$
$$= -0.6987000$$

The result of probability based 10 with n-gram is (-0.6987000) are converted into FST with base log e, so (-0.6987000) value is product into (2.303) value .The product of value with weight was used to construct in Weighted Finite State Transducer in Figure 5.5 with ARPA file based on n-gram such as 2 grams,3 grams, and etc.

Table 5.4 Example of Transitions for 4 Words

From state	To state	Input	Output
0	1	က|◌ာ	k
1	5	လ	a|l
5	7	◌ာ	a
1	4	ဗ	a|b
4	7	◌ာ	a
0	3	က	k
3	7	◌ာ	a
7	10	ဗ|◌ွ	b
10	9	န|◌်	un
0	2	က|ေ◌	k
2	6	◌ာ|◌်	o
6	8	တ|◌ွ	t
8	9	န|◌်	un

In Figure 5.5, Start state is “0” and final states are “7” and “9”. The states between start state and final state are called state. States are “1”, “2”, “3”, “4”, “5”, “6”, ,”7”,“8”, ,”9”,“10”. The symbol “/” followed by value is called weight such as 2.4692, 2.1635 and 2.0457 .etc. The path does not have the transition function is called failure transition such as state “1” occurs the flower transducer that is expressed symbols by the empty string ε or – .

Table 5.4 describes the path of 4 words such as “ကာလာ”, “ကာဗာ”, “ကာဗွန် ”and “ကော်တွန်”. As an Example , path of grapheme “ကာလာ” are 0,1,5 .From the start state 0 to 1 path that generated the output “k” phoneme symbols for input grapheme “က, ◌ာ ” and the next state from state 1 to state 5 path that generated the output ‘a,l’ phoneme symbols for the input grapheme “လ” and final state from 5 to 7 path that generated “a” for the input grapheme “◌ာ” .

5.5 Decoding

Decoding is created a composition phoneme lattice with a single shortest path through the input word. The proposed WFST-based approach is intent to decode using default decoder and the decoding process is summarized in Equation

$$H_{best} = \text{ShortestPath}(\text{Min}(\text{Det}(\text{Project}_o\,(w \circ M)))) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(5.2)$$

where,‘H_{best}’ refers to the best pronunciation hypothesis given the model. ‘w’ is a linear FSA state with the target word, and ‘M’ is a WFST-based representation of the joint n-gram model**.** The ‘∘’ operator refers to weighted composition, which is used to hierarchically cascade multiple WFST. $\text{project}_o(\cdot)$ refers to projecting the output labels (phonemes result) that produces a phoneme word of initial pronunciations.

Composition combines transducers at different levels. For example if F is a finite state grammar and P is a pronunciation dictionary then P∘F transduces a phoneme string to word strings allowed by the grammar. Projection that is converted WFST into a WFSA representing the input language or the output language of the machine. The Det operator refers to an optional determinization step that input sequence has a single path and reduces multiple paths from the input data.

Finally, shortest path (·) refers to the shortest-path algorithm that is used to compute the shortest path across the input machine starting from the start state to a final state. The shortest distance operation is used to compute the shortest distance from the start state to every other state. This process occurs during both the alignment and decoding stages.

5.6 Evaluation

The evaluation of the G2P system is measured by phoneme error rate (PER). The following formula is used to compute PER,

$$PER= (I+D+S)*100/N_P \qquad \ldots\ldots\ldots\ldots\ldots(5.3)$$

where, I=the numbers of insertion

D=the numbers of deletion

S=the numbers of substitution

N_P=the numbers of phonemes in the reference

Phoneme Error Rate (PER) is calculated by dividing the total number of insertions, substitutions, deletions in the hypothesis text by the total number of word in the reference text. And, the result is multiplied by one hundred. SCTK [22] scoring toolkit was set up to evaluate the accuracy of the system in term of Phoneme Error Rate (PER).The example of hypothesis text, reference text and Phoneme Error Rate is depicted as follow;

References for "ဟိုက်ဒရို ကာဗွန်ဒိုင်အောက်ဆိုက်"

h	ai’	d	a-	r	ou	k	a	b	un	d	ain	au'	hs	ai’

Hypothesis for "ဟိုက်ဒရို ကာဗွန်ဒိုင်အောက်ဆိုက်"

h	ai’	d	a-	r	ou	k	a-	b	un	d	ain	au'	hs	ai’

Substitution

Substitutions (S) = 1,

Calculation of PER for word "ဟိုက်ဒရို ကာဘွန်ဒိုင်အောက်ဆိုက်"

PER= (1+0+0)*100/ 15

$=\frac{100}{15}$

=6.67%

CHAPTER 6

SYSTEM IMPLEMENTATION AND EVALUATION

This chapter describes the system implementation, evaluation, 10-fold cross validation, experimental results in details and error analysis. The system is written in the Python programming languages, datasets are retrieved from foreign pronunciation dictionary.

6.1 System Implementation

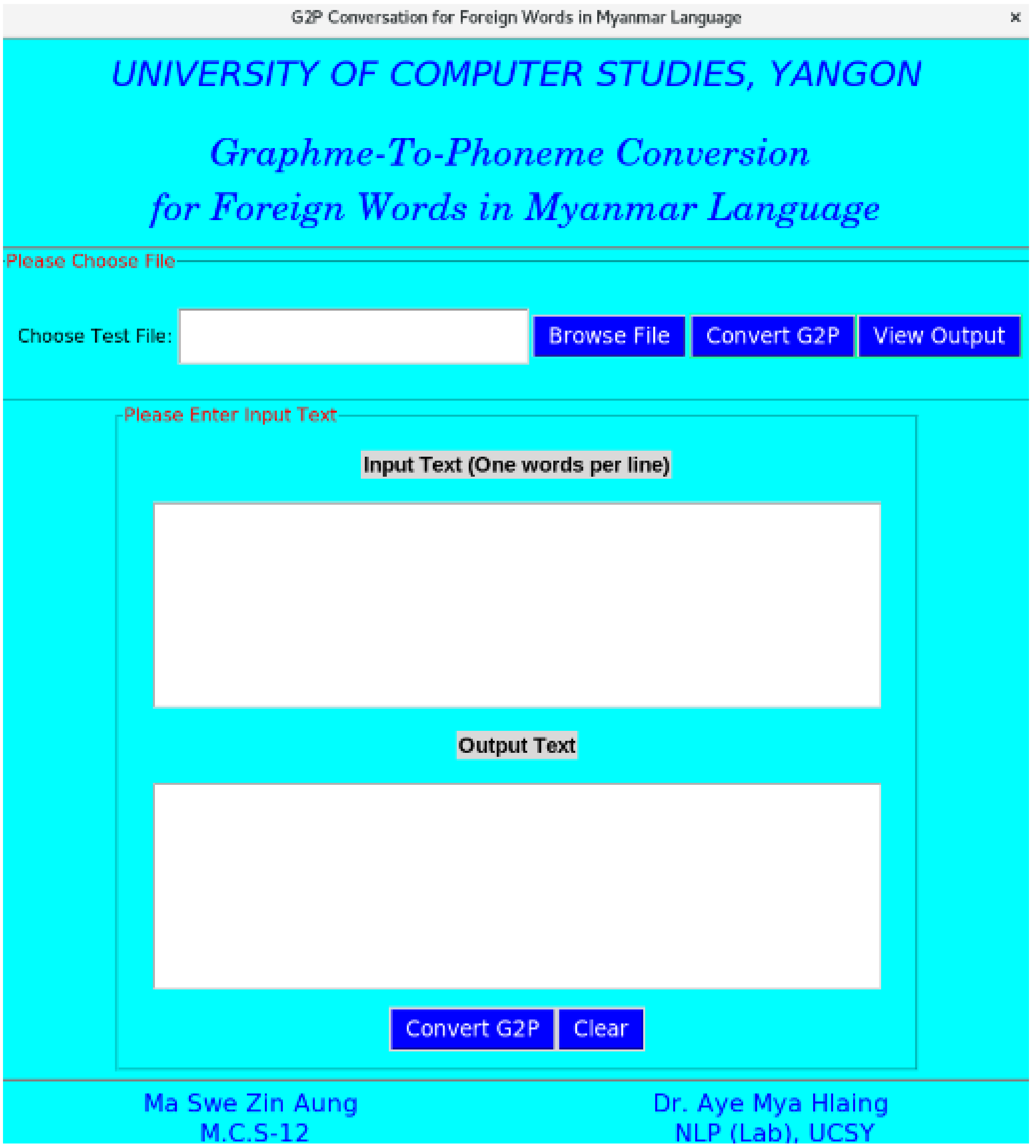

Figure 6.1 Home Page of the System

Figure 6.1 shows the user interface of the system "Grapheme-To-Phoneme Conversion for Foreign Words in Myanmar Language.

The body of the testing is the main part of the system. This place will make testing for user from various types of ways such as choose testing file from “Browse File” button and set input testing text from input textbox. The web page was composed with labels, text box and button.

In Figure 6.2, if the user choose testing file anywhere from the browse button, the “Convert G2P” button convert from the input file with the foreign words into the output file that their pronunciation. When the user clicks “View Output” button, the result file are shown output file. This file contains testing file of pronunciation. Figure 6.3 shows “View Output” file results.

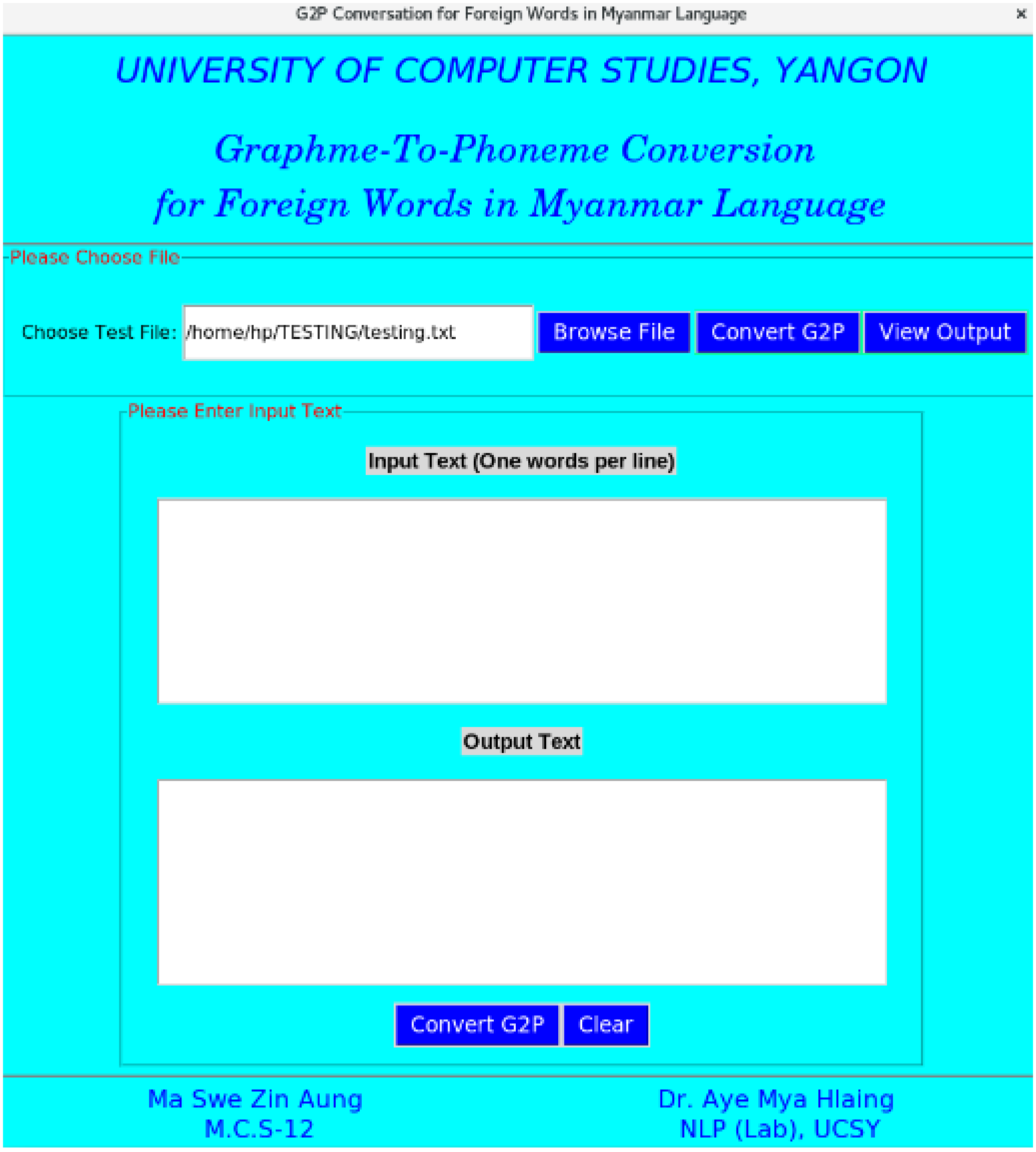

Figure 6.2 Testing with File

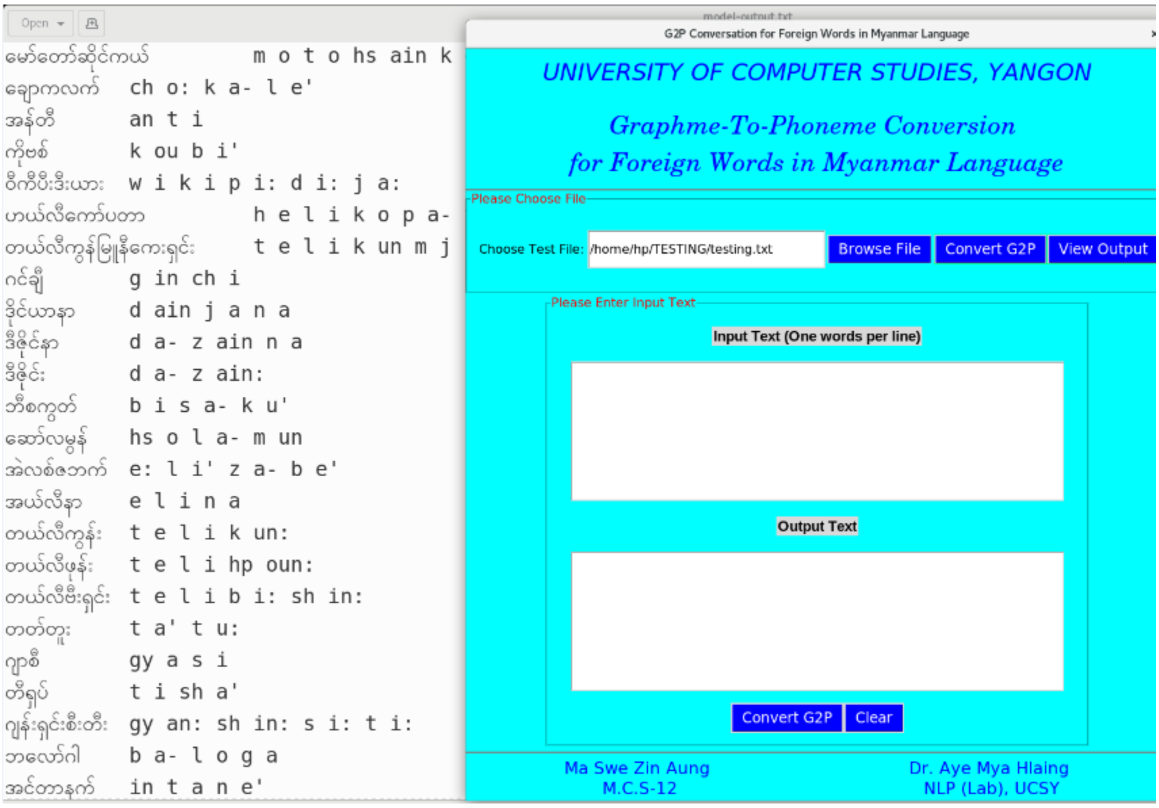

Figure 6.3 View Output File

In Figure 6.4, next ways for testing is set the testing text into input text box where we must carefully set input words that must one word per line. After user set the input testing words, user must click “Convert G2P Button”. If user clicks this button, these input words were changed into their pronunciation as result words. The output pronunciation result was shown in “output text box”. “Clear” button will make clear the text from both input text box and output text box because this button intends to ready testing for another times.

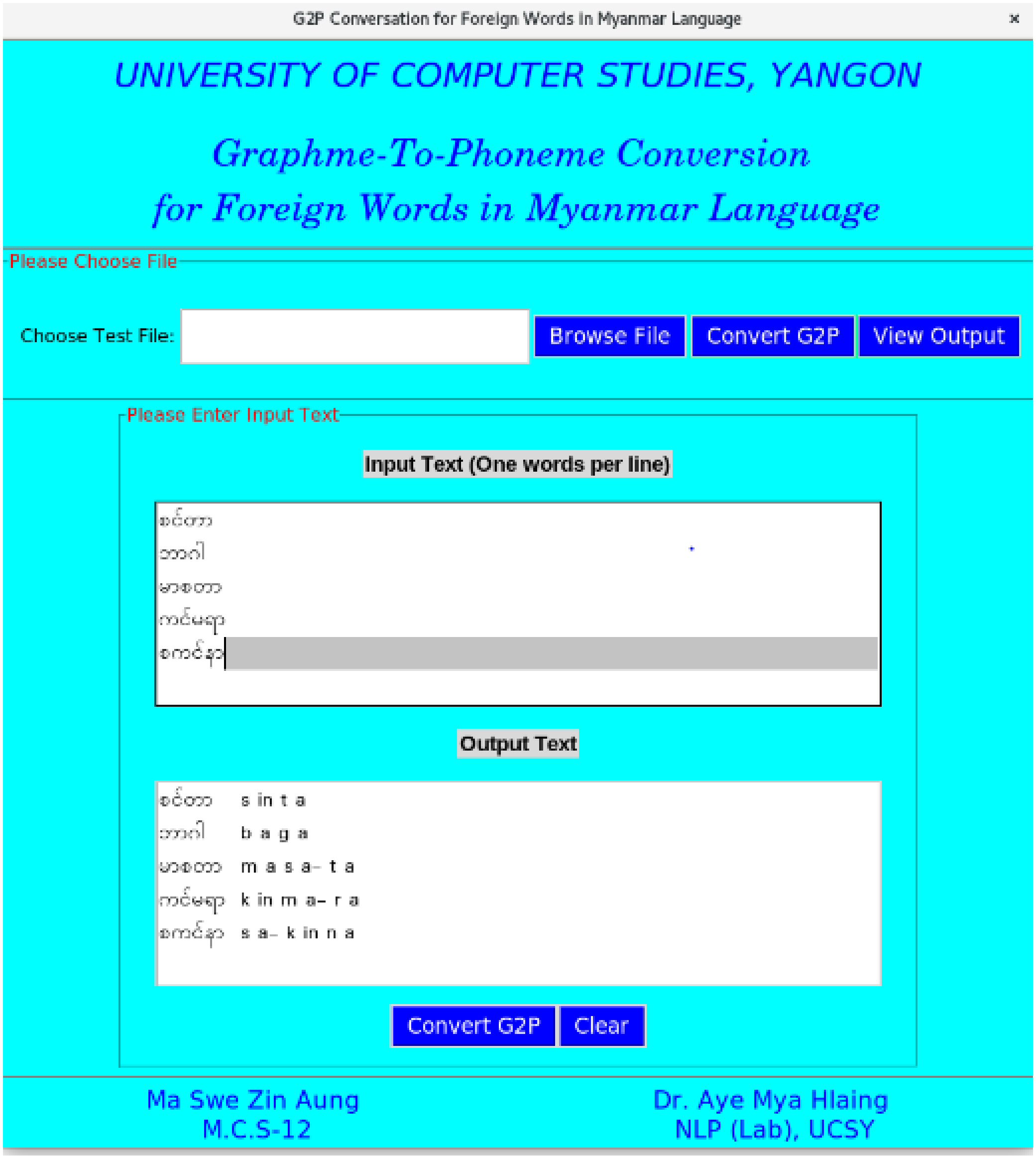

Figure 6.4 Testing with Input Text

6.2 10-Fold Cross Validation

Cross Validation is a re-sampling procedure used to evaluate the performance of model on a limited data sample. A single parameter k in the procedure that means the number of split groups in a given data. This procedure is often called k-fold cross-validation. When a specific number for k is selected, that can be used in hold of *k* -value to the model, for example, *k* value is 10.This situation is called 10- fold cross-validation

The Data is randomly split into two parts that training and testing in this approach. This approach is intended to perform the model training on training data set and the left one is used to testing data set for validation. It is mainly used based on a machine learning model to estimate unseen data. This model is applied on such as Under-fitting/Over-fitting/Well generalized.

The procedure is followed for the *k* "folds":

- A model is trained with k-1 value in order to determine the number of fold as training data;
- The remaining model is validated on the left part of the testing data set that it is used as a test set to compute a performance measure of accuracy

The performance of the model measure by *k*-fold cross-validation is to compute the average numbers of the fold in the loop. This process is repeated 10 times until all permutations are used for training and testing. 10-fold cross validation provides a measure of ensuring the validity of manually prepared pronunciation dictionary and the consistency of the performance of the G2P model.

Data can be divided into two partitioning on evaluations within most NLP settings. Training data- data set where input data are paired with the desired outputs, regard as the result of manual symbol. Test data - describes the data that will be used to evaluate the performance of the system after validation has been done.

6.3 Evaluation Results

The size of training data and testing data are shown in Table 6.1.The results are shown in Tables 6.2, PER was obtained 2.36 % in average section shows how to calculate the algorithm of a text classification system step by step using a small dataset as an example.

Table 6.1 Data Set

Data Set	Number of Foreign Words
Training Data	30,600
Testing Data	3,400

Table 6.2 Experimental Results

Models	PER (%)
1	2.8%
2	1.6%
3	2.1%
4	1.6%
5	1.9%
6	2.8%
7	2.5%
8	2.7%
9	2.8%
10	2.8%
Average	**2.36%**

In Table 6.2, performance of ten models result were balance accuracy ensure the validity of manually prepared pronunciation dictionary and the consistency of the performance of the model.

6.4. Error Analysis

Error Analysis is based on the hypothesis text. The following errors were found in G2P conversion model. In Table 6.3, foreign word "ဂျီအန်စု" of its pronunciation is "gy i an s u." in the reference text but the model was generated "gy i an z u." that mean "ဂျီအန်ဇု". This error also is "Consonant Error".

Table 6.3 Example of Some Errors

Foreign words	Reference	Hypothesis	Error Type
ဂျီအန်စု	gy i an s u.	gy an z u.	Consonant Error
မက္ကဆီကို	m e' k a- hs i k ou	m e' k a- hs i g ou	Consonant Error
ဂျော့ချ်အိုလီဗာ	gy o. CH ou l i b a	gy o. a CH ou l i b a	Vowel Error
ဘရောင်း	b a- j aun:	b a- r aun:	Vowel Error
ဂျော	gy o:	gy o.	Vowel Error
ဟာလာဟူစိန်	h a l a h u s ein	h a- l a h u s ein	Vowel Error
ဂျွန်အာဖ်	gy un a HP	gy un a H P	Foreign Error
ဂျွန်လော့ခ်	gy un l o. KH	gy un l o. kh	Foreign Error

The second error is “Vowel Error”. For example of vowel error, “ဂျော” of its pronunciation is “gy o:” but the model was generated “gy o.” that mean “ဂျော့”, so, this error was vowel error.

The next error is “Foreign Error”. For example of foreign error, “ဂျွန်အာဖ်” of its pronunciation is “gy un a HP”. But the model was generated “gy un a H P”, where the instead of “HP”, model was substitute the phoneme symbols “H P”. So, this error was foreign error.

CHAPTER 7
CONCLUSION

This chapter summarizes the research work of G2P conversion for foreign word in Myanmar Language and presents the advantages and limitations of the system. Furthermore, this chapter also describes future work on G2P model.

7.1 Thesis Summary

This thesis presented the research results of WFST based on N-gram language model with G2P conversion for the foreign words in Myanmar Language. This thesis is composed of seven chapters and Chapter 1 described the objective and G2P conversion and their related works.

Chapter 2 describes about the Myanmar Language and the phoneme symbols used for consonants, vowels, tones, medial and foreign words. The basics of n-gram modeling, introduce the smoothing and their type, Kneser-Ney smoothing ,WFST and FST were presented in Chapter 3.Theses techniques were applied in G2P conversion model .This chapter expressed usage and equation for each techniques.

Chapter 4 depicts the process of building pronunciation dictionary for foreign words that compose how to prepare for training data and the statistics of phoneme symbols in the final pronunciation dictionary. Chapter 5 describes G2P modeling of the purposed system that include the process of alignment, training the model of n-gram language model, how to convert from ARPA with joint-sequence n-gram into WST model, how to decode the input grapheme to their pronunciation. It also expressed about evaluation with the example for a word based on both reference and hypothesis.

Chapter 6 presented about the system implementation of the model and 10-fold cross validation that ensure the validity of manually prepared pronunciation dictionary and the consistency of the performance of the G2P model. The experimental result of training and testing with WFST base model is 2.36% in average Phoneme Error Rate (PER).

In this thesis, a WFST based G2P model for automatic G2P conversion of foreign words in Myanmar language was developed. It is an important step for Myanmar ASR and TTS development. The pronunciation dictionary of foreign words was manually built and it consists of 34,000 entries. Joint N-gram language modeling and Weighted Finite State Transducer (WFST) based approach has been applied in modeling the G2P conversion. 10-fold cross validation is done on pronunciation dictionary and it got 2.36% in average

Phoneme Error Rate (PER). This G2P conversion model can be applied in many application areas such as Text-to-Speech (TTS) and Automatic Speech Recognition (ASR) in Myanmar language, and the pronunciation dictionary can be directly utilized in TTS and ASR application.

7.2 Limitation of the System

The G2P System can be applied in many application areas such as text-to-speech (TTS) and automatic speech recognition (ASR). The system was built by using pronunciation dictionary of foreign words in Myanmar language so it can generate phoneme symbols for foreign words. It cannot correctly generate the pronunciation of normal Myanmar pronunciation because the G2P model was trained based on foreign words.

7.3 Future Work

In future work, pronunciation dictionary will be extended for wide coverage and different Grapheme-to-Phoneme Conversion techniques such as sequence to sequence model, conditional random field, support vector machine based point-wise classification will be applied on the dictionary. Other Language modeling techniques such as Recurrent Neutral Network Language Model (RNNLM), Stanford Research Institute Language Modeling Toolkit (SRILM) can be applied in building language model.

AUTHOR'S PUBLICATIONS

[1] Swe Zin Aung, Aye Mya Hlaing, "Grapheme-To-Phoneme Conversion for Foreign Words in Myanmar", National Journal of Parallel and Soft Computing, University of Computer Studies, Yangon, Myanmar, 2022.

REFERENCES

[1] A. M. Hlaing, W. P. Pa "Sequence-to-Sequence Models for Grapheme to Phoneme Conversion on Large Myanmar Pronunciation Dictionary" (IEEE)*O-COCOSDA 2019: 1-5*

[2] A. M. Hlaing ,"Ehhancing Myanmar Text-To –Speech System by using Liguistic Information LSTM-RNN Based Speech Synthesis Model And Text Normalization",2020

[3] Allauzen and M. Riley and J. Schalkwyk and W. Skut and M. Mohri. OpenFST: A General and Efficient Weighted Finite-State Transducer Library, Proc. CIAA 2007, pp. 11-23.

[4] Bo-June (Paul) Hsu ,"Language Modeling for Limited-Data Domains", 2009

[5] B. M. Cartney," NLP Lunch Tutorial: Smoothing",2005

[6] D. Caseiro, I. Trancoso, L. Oliveira, and C. Viana , " Grapheme to-phone using finite-state transducers, " in In: Proc. 2002 IEEE Workshop on Speech Synthesis. Volume, 2002, pp. 1349–1360.

[7] Department of the Myanmar Language Commission Myanmar-English Dictionary, Yangon, Ministry of Education, 1993.

[8] Hsu and J. Glass, "Iterative Language Model Estimation: Efficient Data Structure & Algorithms", Proc. Interspeech 2008.

[9] J. R. Novak, N. Minematsu and K. Hirose,"Phonetisaurus: Exploring grapheme-to-phoneme conversion with joint n-gram models in the WFST framework", Natural Language Engineering, 22(6), 2016, pp.907-938.

[10] Kneser, R., and Ney, "Improved backing-off for m-gram language modeling. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, , Detroit, Michigan, 1995,pp. 1:181–4.

[11] M. Bisani, H. Ney," Joint-sequence models for grapheme-to-phoneme conversion", Speech Communication50, 2008, p p. 434-451.

[12] M.Korner, "Implementation of Modified Kneser-Ney Smoothing on Top of Generalized Language Models for Next Word Prediction" ,2013

[13] Pearson "Speech and Language Processing: International Edition",2008, ISBN 10: 0135041961 ISBN 13: 9780135041963

[14] Soky, Kak, X. Lu, P. Shen, H. Kato, H. Kawai, C. Vanna, and V. Chea.(2016) "Building WFST based Grapheme to Phoneme Conversion for Khmer", in Proceedings of KNLP2016.

[15] S. Jiampojamarn, et.al, "Applying Many-to-Many Alignments and Hidden Markov Models to Letter to-Phoneme Conversion", NAACL HLT, pp.372379, 2007.

[16] S. Jiampojamarn, G. Kondrak, "Letter-to-Phoneme Alignment: an Exploration", Proc. ACL, pp.780788, 2010.

[17] Y. K. Thu, W. P. Pa, Finch Andrew, A. M Hlaing, H. M. S.Naing, S. Eiichiro, and H. Chiori, "Syllable Pronunciation Features for Myanmar Grapheme to Phoneme Conversion," in the 13th International Conference on Computer applications (ICCA),Yangon, Myanmar, February, 2015,pp.161-167.)

[18] Y.K. Thu, W. P. Pa, Y. Sagisaka, and N. Iwahashi, "Comparison of Grapheme-to-Phoneme Conversion Methods on a Myanmar Pronunciation Dictionary," Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing ,Osaka, Japan, December, 2016,pp.11-22.

[19] Y.K .Thu,W. P. Pa, A.Finch, J.Ni, E. Sumita, and C. Hori" The Application of Phrase Based Statistical Machine Translation Techniques to Myanmar Grapheme to Phoneme Conversion",2015

[20] https://www2.nict.go.jp/astrecatt/member/mutiyama/ALT

[21] https://en.wikipedia.org/wiki/Brahmi_script

[22] https://github.com/usnistgov/SCTK/.