Grapheme-to-IPA Phoneme Conversion for Burmese (myG2P Version 2.0)

Honey Htun, Ni Htwe Aung, Shwe Sin Moe, Wint Theingi Zaw, Nyein Nyein Oo, Thepchai Supnithi, Ye Kyaw Thu

(Honey Htun, Ni Htwe Aung, Shwe Sin Moe, and Wint Theingi Zaw equally contributed to this work as first authors.)

***Abstract*— Grapheme-to-Phoneme (G2P) conversion is the task of generating the pronunciation on a given word in the form of written text. This plays an important role in the field of automatic speech recognition (ASR) and speech synthesis. There are lack of Grapheme-to-IPA phoneme dictionaries for low-resourced languages, especially, Burmese (Myanmar language). In this paper, we introduced the Grapheme-to-IPA phoneme pairs for Burmese these are manually prepared and based on myG2P dictionary (version 1.1). The conversion models are also developed by applying five methods and they are Ripple Down Rules (RDR), Hidden Markov Model (HMM), Conditional Random Field (CRF), Phrase-based Statistical Machine Translation (PBSMT) and Bi-directional Long Short-Term Memory (Bi-LSTM), one type of Recurrent Neural Networks (RNN). Moreover, we evaluated the proposed conversion models based on the accuracy, F1-score and phoneme error rate (PER).**

***Index Terms*—Grapheme-to-IPA phoneme conversion, Burmese (Myanmar), RDR, HMM, CRF, PBSMT, Bi-LSTM (RNN).**

I. Introduction

THE process of grapheme-to-phoneme (G2P) conversion process is critical in ASR and Text-to-Speech (TTS) research fields. Generating the pronunication of a word is the middle-layer between the acoustic and language models and the quality of pronunciations will reflect the performance of ASR and TTS systems. Converting to the phonetic symbol just by looking at the grapheme is practically difficult. This is because the actual pronunciations may differ based on the adjacent contexts in some cases. To get the correct pronunciation of spoken words or words in the written text, we need to have the grapheme-to-phoneme dictionary. This type of dictionary can solve the encountered problems and make sure to get the good performance in ASR and TTS systems. However, such dictionaries are extremely rare for low-resourced languages, especially for Burmese (Myanmar language). Motivated by this, we developed "Grapheme-to-IPA" (G2IPA) dictionary for Burmese as one main contribution and did experiments by applying RDR, HMM, CRF, PBSMT, and Bi-LSTM (RNN) approaches, respectively. We used three metrics such as accuracy, F1, and phoneme error rate (PER) to measure the experimental results and made the error analysis in detail. We expect the proposed G2IPA dictionary and those studies would be the applicable investigations in ASR and TTS systems for Burmese.

We describe the previous works of G2P conversion systems in the next section. Section III introduces the nature of Myanmar language and section IV describes about the process of G2IPA conversion, our main contribution, for Burmese. In section V, the IPA phoneme symbols corpus building is presented. Then, we describe the experimental methodologies used in G2IPA conversion experiments in section VI. The experimental settings are presented in section VII. The experimental results are reported with discussions in section VIII. In section IX, we analyze some of the errors encountered on the converted outputs. The final section describes the conclusion and future work.

II. Related Work

The previous works of some G2P conversion systems (including Burmese) are presented in this section.

Ei Phyu Phyu Soe [1] proposed the first dictionary-based approach for Myanmar G2P conversion. In this system, only Myanmar syllables were analyzed although the dictionary consists of Pali or subscript consonants. There were some out of vocabulary (OOV) errors due to the dictionary-based approach.

Ye Kyaw Thu et al. [2] investigated the first G2P mappings for Burmese. In this paper, four simple pronunciation patterns were proposed that can be used as basis features. These patterns were used in the baseline CRF approach for G2P conversion. The experiments are done based on the two levels: word and phoneme. The results

Honey Htun, Ni Htwe Aung, Shwe Sin Moe, Wint Theingi Zaw and Nyein Nyein Oo are with the Department of Computer Engineering and Information Technology, Yangon Technological University, Insein, Yangon, 11011 Myanmar, Corresponding author email: honeyhtun.85@gmail.com, nhadec@gmail.com,shwesinmoe.ytu25@gmail.com, wint.wtgz@gmail.com and nno2005@gmail.com.

Thepchai Spnithi is with LST Lab., NECTEC, Thailand, Corresponding author emal: thepchai.supnithi@nectec.or.th.

Ye Kyaw Thu is with LST Lab., NECTEC, Thailand, Corresponding author email: yktnlp@gmail.com.

Manuscript received February 8, 2021; accepted March 19, 2021; revised April 24, 2021; published online April 26, 2021

showed that the accuracies of both levels were increased by adding these new patterns to the baseline CRF model.

Ye Kyaw Thu et al. [3] investigated Myanmar G2P conversion system using phrase-based statistical machine translation (PBSMT) approach. We found that PBSMT achieved the higher performance than CRF and can predict the new pronunciations on the new compound words. This approach can also deal with the influence of surrounding words on the pronunciation of a word.

Ye Kyaw Thu et al. [4] analyzed the performance of seven G2P conversion approaches such as Adaptive Regularization of Weight Vectors (AROW) based structured learning (S-AROW), 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. The outputs are measured by using automatic evaluation of PER and made manual checking. The experimental results showed that CRF, PBSMT and WFST approaches can improve the performance of G2P conversion systems for Burmese.

Kanishka et al. [5] proposed the unidirectional LSTM (ULSTM) with different kinds of output delays and deep bidirectional LSTM (DBLSTM) with a connectionist temporal classification (CTC) layer. The experimental results showed that the DBLSTM-CTC model achieved word error rate (WER) of 25.8% on the public CMU dataset for United State (US) English. And, combining the DBLSTM-CTC model with a joint n-gram model gained in WER of 21.3%, which is a 9% relative improvement, compared to the previous best WER of 23.4% from a hybrid system.

Based on the knowledge of the previous works, the experiments are carried out to study the performance of G2IPA and "IPA-to-Grapheme" (IPA2G) conversions for Burmese in this paper.

III. Nature of Burmese (Myanmar Language)

Burmese is a tonal language and a member of the Lolo-Burmese branch of the Sino-Tibetan family. Burmese script supported Brahmi script which developed in India from about 500 B.C. to over 300 A.D. It's spoken mainly in Myanmar (Burma), because it is the official language. In 2007, there have been about 33 million the Bama (Burman) people who used Burmese as a primary language. There are also another 10 million, particularly ethnic minorities in Myanmar and neighboring countries, who speak it as a second language. The Burmese consists of 12 vowels, 33 consonants, and 4 medials which are used as basic alphabets [6]. The Burmese is also syllable-based. It uses SOV (Subject-Object-Verb) typology and a Myanmar text has to be written in order from left to right. It does not have explicit word boundary markup (regular inter-word spacing) between words, although modern writing usually contains spaces after each clause to enhance readability. Phrases are separated by using space, rather than words.

We need to know the nature of Burmese before the process of G2IPA phoneme conversion. The following subsections introduced the basic components of Burmese (Myanmar language) such as consonants, vowels, and types of tone.

A. Consonants

There are 33 basic letters to indicate the initial consonant of a syllable and four diacritics to show the additional consonants in the onset. These are known as "Byee" in Burmese. Consonants are used as the base characters of Myanmar words, and they are similar in pronunciation to other Southeast Asian scripts such as Thai, Lao and Khmer. Medials are known as "Byee Twe" in Burmese. There are 4 basic medials and 6 combined medials in the Burmese script [8].

B. Vowels

There are 12 basic vowels such as 8 monophthongs and 4 diphthongs in Burmese script. Table I shows Myanmar monophthongs and diphthongs which are classified consistent with tongue heights and positions. These 12 vowels are presented by diacritics, which are placed above, below, before or after the consonant character [9].

TABLE I: Myanmar Vowels

Tongue Height	**Monophthongs**		**Diphthongs**	
	Front	**Back**	**Front offglide**	**Back offglide**
Close	i	u		
Close-mid	e	o	ei	ou
Mid	ə			
Open-mid	ɛ	ɔ		
Open		a	ai	au

C. Tones

Burmese is a tonal language, which suggests phonemic differences are made on the basic of the tone of a vowel. A normal syllable structure consists of an initial consonant which is followed by a vowel with an associated tone. This indicates all syllables in Myanmar have prosodic features. Different tone makes different meanings for syllables with the identical structure of phonemes. In the Myanmar writing system, a tone is indicated by a diacritic mark. There are four tones in Burmese namely low, high, creaky and stopped (checked) [10]. Table II shows the phonemic transcriptions for making tones, using "က" /ka/ syllable as the base example.

TABLE II: Example of Four Myanmar Tones

Tone Name	**Myanmar Word**	**IPA Symbol**	**Description in English**
Creaky	က	ka̰	Dance
Low	ကာ	kà	Protect
High	ကား	ká	Car
Stopped(Checked)	ကပ်	kaʔ	Stick

IV. Grapheme to IPA Phoneme Conversion for Burmese

This section illustrates our main contribution, the development of myG2P (version 2.0) dictionary, with some points to consider and examples. This development process uses the syllable onset, syllable rhymes, and phonetic transcriptions in [11] as the basic for IPA phonemic tagging to G2P pairs in myG2P (version 1.1). This phonetic dictionary of myG2P (version 1.1) was based on the Myanmar Language Commission (MLC) dictionary [12]. In myG2P (version 1.1), these original MLC dictionary was extended with additional phonetic symbols for foreign words to get the closet and accurate pronunciation symbols [2].

Myanmar 33 consonants are represented by 23 phonemes since some consonantal letters represents an equivalent phoneme, for instance, the consonants /ဂ/ and /ဃ/ represent the same phoneme/ga̤/, the consonant /ဒ/ and /ဓ/ represent the same phoneme /da̤/. Table III and IV provide the alphabet, the syllable onset in IPA and the way the alphabet is referred to in Burmese. These may be either a descriptive name or just the sound of the alphabet which are arranged in the traditional order [7]. The IPA symbols for Myanmar consonants are classified based on the place of articulation and the manner of articulation. The pronunciation of vowel graphemes may vary in open and closed syllables. Table V shows 7 independent vowels that are commonly used for Pali or Sanskrit words [11].

Burmese has only one prescript vowel-sign "ေ◌" (presented in IPA by /è/) that appears to the left of the base consonant letter, for example, "ဖေဖေ" /pʰèpʰè/ ('father' in English). The standalone vowel sounds are generally presented using vowel-signs applied to "အ", for example, "အိတ်" /ʔeɪʔ/ ('bag' in English). In this example, "အ" is described by only one IPA symbol /ʔ/ instead of the original symbol /ʔə/. As the next standalone vowel example, "အတန်း " ('class' in English) is pronunced by /ʔətáɴ/. Its standard IPA symbol /ʔə/ can only be used because "အ" is not combined with other vowel signs.

The relationship between the word and its pronunciation is not simple and this can be ambiguous depending on the neighbouring contexts, Part-Of-Speech (POS), and so on. Some Myanmar syllables can be pronounced as their written forms without pronunication changes, for example, three-syllable word "ကယ်တင်ရှင်" ('savior' in English) is pronounced as the standard pronunciation of each syllable (/kɛ̀ + tɪ̀ɴ + ʃɪɴ ==> kɛ̀tɪ̀ɴʃɪɴ/). However, POS plays an important role in pronunciation; consider the following Myanmar word "စာရေး" can be pronounced in two ways: /sàjé/ when used as **verb** ("write" in English) and /səjé/ when used as **noun** ("clerk" in English). There is one point we should consider in pronunciation, known as the vocalic weakening process, that can affect the first syllables of certain words (mostly nouns and adverbs), for example, "ကုလားထိုင်" is pronounced as /kələtʰàɪɴ/ ('chair' in English), not /kṳlátʰàɪɴ/ and "ဘုရား" is pronounced as /phəjá/ ('pagoda' in English), not /bṳjá/.

Some Myanmar words with different meanings can have the same pronunciations, for example, both " ကတိ " ("promise" in English) and " ဂတိ " ("next existence" in English) can be pronounced as /gədi̤/. Local dialect can affect the pronunciations of some syllables such as Myanmar word " ချော်ရည် " ("lava" in English) can have both /tɕʰɔ̀jè/ and /tɕʰɔ̀jì/ pronunciations, " ချွေးစေး " ("sweat" in English) can be pronounced as /tɕʰwézé/ and /tɕʰwézí/.

There is one point to consider in some exceptional cases: some Myanmar syllables with Kinzi have to be pronunced by adding high tone " ◌း " ("visarga") symbol such as /gíɴgà/ for " ဂင်္ဂါ " (the name of river), /θíɴbɔ́/ for " သင်္ဘော " ("ship" in English) and so on. Most of the pronunciation changes can occur from unvoiced to voiced, for example, the change from: /pi̤tóʊɴ/ to /bədóʊɴ/ for " ပိတုန်း " ("bumblebee" in English), /ta̤páʊɴ/ to /dəbáʊɴ/ for " တပေါင်း " (the first month of Myanmar calendar). Burmese vowel-signs are all combining characters and examples of IPA labels for "Ka" consonant with vowel combinations are shown in Table VI.

V. Corpus Building for IPA Phoneme Symbols

We built the G2IPA phonetic dictionary, used for training the G2IPA phoneme conversion models, based on myG2P (version 1.1) [2] in the way of adding IPA symbols. This myG2P dictionary (version 1.1) based on the MLC dictionary and was used for VoiceTra (Multilingual Speech Translation Application) Myanmar language project of NICT, Japan (during 2014-2015). In myG2P (version 1.1), Burmese (Myanmar) words of MLC dictionary are segmented into syllable units by using Regular Expression based Myanmar Syllable Segmentation Tool named "sylbreak" [13]. The phonemic mappings for foreign words are also extended and some of the phoneme mappings are modified to ensure the consistency of phonemic order.

We modified this myG2P (version 1.1) by mapping its syllable Burmese (Myanmar) words and their phonetics to the proposed IPA symbols based on the syllable onset in Table (III, IV, V, and VI), syllable rhymes, and phonetic transcriptions in [11] as the references. There are 2,353 unique syllables, 1,928 unique IPA symbols and a total of 24,803 G2IPA pairs in our current myG2P (version 2.0). These pairs were prepared and checked many times manually by two Ph.D. candidates and two master candidates who are from the department of Computer Engineering and Information Technology, Yangon Technological University, Myanmar. Some example pairs are presented in Table (VII) as follows.

VI. Methodology

In this section, we describe the methodologies that are applied in the IPA tagging process of Burmese words in myG2P (version 2.0) dictionary.

A. Ripple down rules (RDR)

RDR is an incremental construction method and provides a structure for rule-based classifiers [14]. it is orga-

TABLE III: IPA Symbols for Group Consonants

Group Name	Unaspirated (သိထိလ)		Aspirated (ဓနိတ)		Voiced (လဟု)				Nasal (နိဂ္ဂဟိတ)	
Velars (ကဏ္ဍဇ) ကဝဂ်	က	/k/	ခ	/kʰ/	ဂ	/g/	ဃ	/gʰ/	င	/ŋ/
	ကကြီး [ka̰ dʑí]		ခခွေး [kʰa̰ gwé]		ဂငယ် [ga̰ ŋɛ̀]		ဃကြီး [ga̰ dʑí]		င [ŋa̰]	
Palatals (တာလုဇ) စဝ	စ	/s/	ဆ	/sʰ/	ဇ	/z/	ဈ	/zʰ/	ဉ /ည	/ɲ/
	စလုံး [sa̰ lóʊɴ]		ဆလိမ် [sʰa̰ lèɪɴ]		ဇကွဲ [za̰ gwɛ́]		ဈမျဉ်းဆွဲ [za̰ mjìɴ zwɛ́]		ညကလေး ညကြီး [ɲa̰ dʑí]	
Alveolars (မုဒ္ဓဇ) ဋဝဂ်	ဋ	/t/	ဌ	/tʰ/	ဍ	/d/	ဎ	/dʰ/	ဏ	/n/
	ဋသန်လျင်းချိတ် [ta̰ təlín dʑeɪʔ]		ဌဝမ်းဘဲ [tʰa̰ wʊ́ɴ bɛ́]		ဍရင်ကောက် [da̰ jìɴ gaʊʔ]		ဎရေမှုတ် [da̰ jè m̥oʊʔ]		ဏကြီး [na̰ dʑí]	
Dentals (ဒန္တဇ) တဝဂ်	တ	/t/	ထ	/tʰ/	ဒ	/d/	ဓ	/dʰ/	န	/n/
	တဝမ်းပူ [ta̰ wʊ́ɴ bù]		ထဆင်ထူး [tʰa̰ sʰìɴ dú]		ဒထွေး [da̰ dwé]		ဓအောက်ခြိုက် [da̰ ʔaʊʔ tɕʰaɪʔ]		နငယ် [na̰ ŋɛ̀]	
Labials (ဩဋ္ဌဇ) ပဝဂ်	ပ	/p/	ဖ	/pʰ/	ဗ	/b/	ဘ	/bʰ/	မ	/m/
	ပစောက် [pa̰ zaʊʔ]		ဖဦးထုပ် [pʰa̰ ʔóʊʔ tʰoʊʔ]		ဗထက်ခြိုက် [ba̰ lɛʔ tɕʰaɪʔ]		ဘကုန်း [ba̰ góʊɴ]		မ [ma̰]	

TABLE IV: IPA Symbols for Miscellaneous Consonants

Group Name	Unaspirated (သိထိလ)		Aspirated (ဓနိတ)		Voiced (လဟု)				Nasal (နိဂ္ဂဟိတ)	
Without group (အဝဂ်)	ယ	/j/	ရ	/j/	လ	/l/	ဝ	/w/	သ	/θ/
	ယပက်လက် [ja̰ pɛʔ lɛʔ]		ရကောက် [ja̰ gaʊʔ]		လငယ် [la̰ ŋɛ̀]		ဝ [wa̰]		သ [θa̰]	
			ဟ	/h/	ဠ	/l/	အ	/ʔ/		
			ဟ [ha̰]		ဠကြီး [la̰ dʑí]		အ [ʔə]			

TABLE V: IPA Symbols for Independent Vowels

Unaspirated (သိထိလ)		Aspirated (ဓနိတ)		Voiced (လဟု)			
ဣ	/ʔḭ/	ဤ	/ʔì/	ဥ	/ʔṵ/	ဦ	/ʔù/
ဧ	/ʔè/	ဩ	/ʔɔ́/	ဪ	/ʔɔ̀/		

TABLE VI: Some IPA labels for vowel combination with a consonant "Ka"

Grapheme	IPA	Grapheme	IPA	Grapheme	IPA	Grapheme	IPA
က	[ka̰],[kə]	ကူး	[kú]	ကိမ်း	[kéɪɴ]	ကောက်	[kaʊʔ]
ကာ	[kà]	ကံ	[kàɴ]	ကိုက်	[kaɪʔ]	ကောင်	[kàʊɴ]
ကား	[ká]	ကုတ်	[koʊʔ]	ကိုင်	[kàɪɴ]	ကောင့်	[ka̰ʊɴ]
ကက်	[kɛʔ]	ကပ်	[kaʔ]	ကိုင့်	[ka̰ɪɴ]	ကောင်း	[káʊɴ]
ကင်	[kìɴ]	ကမ်	[kàɴ]	ကိုင်း	[káɪɴ]	ကော့	[kɔ̰]
ကင့်	[kɪ̰ɴ]	ကမ့်	[ka̰ɴ]	ကို့	[ko̰]	ကော်	[kɔ̀]
ကင်း	[kíɴ]	ကမ်း	[káɴ]	ကွန့်	[kʊ̰ɴ]	ကို	[kò]
ကစ်	[kɪʔ]	ကယ်	[kɛ̀]	ကွန်း	[kʊ́ɴ]	ကိုး	[kó]
ကတ်	[kaʔ]	ကိမ်	[kèɪɴ]	ကွန်	[kʊ̀ɴ]	ကိပ်	[keɪʔ]
ကဲ	[kɛ́]	ကွမ်	[kʊ̀ɴ]	ကုံး	[kóʊɴ]	ကိတ်	[keɪʔ]
ကဲ့	[kɛ̰]	ကေး	[ké]	ကုမ်	[kòʊɴ]	ကွပ်	[kʊʔ]
ကု	[kṵ]	ကေ	[kè]	ကုမ့်	[ko̰ʊɴ]	ကွတ်	[kʊʔ]
ကူ	[kù]	ကေ့	[kḛ]	ကုမ်း	[kóʊɴ]	ကော	[kɔ́]

nized as trees and automatically restructures transformation rules within the sort of one Classification Ripple Down Rules (SCRDR) tree [15] [16]. A case (data to be classified) enters the basis node and ripples down a specific path to receive its classification [14]. Therefore, a SCRDR are often written as if X then Y where X is named condition and Y is mentioned because the conclusion [17].

For example, the SCRDR tree in Fig. 1, given a case "က" /gə/ "စား" /zá/ "ကွင်း " /gwíɴ/ where " က" /gə/ is a first Burmese word and IPA symbol pair. The case satisfies the condition of the rule (0). As the case does not meet the condition in rule (1), it will be changed to rule (2) using if-not link. As the case satisfies rule (2), then it is continued to rule (3) using except link. It also satisfies rule (3). Therefore, the next tag of the current Burmese word is "zá". The second Burmese word is " စား" /zá/

TABLE VII: Examples of Grapheme-to-IPA Pairs in myG2P (version 1.1 and 2.0)

myG2P (version 1.1)	**Myanmar Word**	**myG2P (version 2.0)**
hpjí sei hpjí sei	ဖြစ် စေ ဖြစ် စေ	$\mathrm{p^h}$jɪʔ sè $\mathrm{p^h}$jɪʔ sè
ga- zaː gwé	က စား ကွက်	gə zá gwɛʔ
khí shei. bjei:	ခေတ် ရှေ့ ပြေး	$\mathrm{k^h}$ɪʔ ʃḛ bjé
ga. ga. na. na.	ဂ ဃ န ဏ	ga̰ ga̰ na̰ na̰
ga- na nou. hsun:	ဃ နာ နို့ ဆွမ်း	gə nà no̰ $\mathrm{s^h}$ʊ́ɴ

and then it is continued to rule (4) using except link. It does not satisfy the condition of the rule (4) again and it will be changed to rule (5) using if-not link. Here, the case satisfies the rule (5). Therefore, the next tag of the current Burmese word is "gwɪ́ɴ".

The case in SCRDR is evaluated by passing a case to the root node (Rule 0) in Fig. 1. Rule (1) is the exception rule of the default node (Rule 0). Rule (2) is the if-not child node of the Rule (1) and an exception link of the Rule (0). And then, Rule (6) and (7) are exception rules of Rule (1) etc. [17].

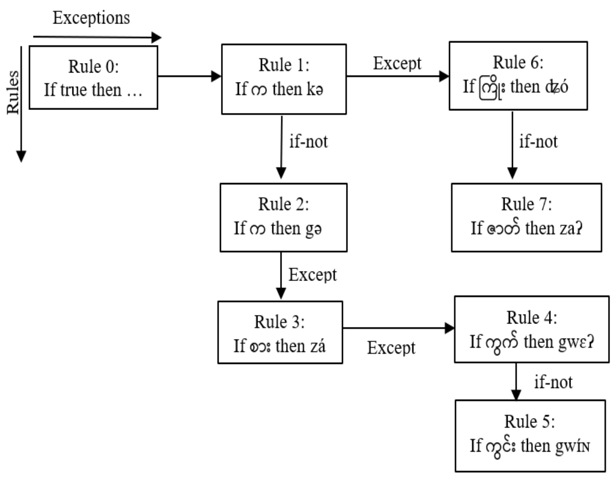

Fig. 1: An example binary tree of Single Classification Ripple Down rules

B. Hidden Markov model (HMM)

HMM may be a probabilistic system designed to model a sequence as a result of a markovian process that can't be observed. The sequence is generated by two stochastic processes: the transition between states and the emission of a letter from each state [18]. It computes a joint probability distribution over possible sequences of labels and chooses the best label sequence [19], [20]. The model describes the joint state and observation sequence as in (1).

$$
\begin{aligned}
&p(y_1,...,y_n,x_1,...,x_n) \\
&= p(y_1)(p(x_1|y_1)\prod_{i=2}^{n} p(y_i|y_{i-1})p(x_i|y_i)
\end{aligned} \tag{1}
$$

Figure 2 shows the state sequence variables $Y_1, Y_2, ..., Y_n$ is not observed (i.e. hidden) and the $X_1, X_2, ..., X_n$ are observable variables (i.e. output) of the Markov chain. Horizontal and vertical arrows indicate transition probability and emission probability.

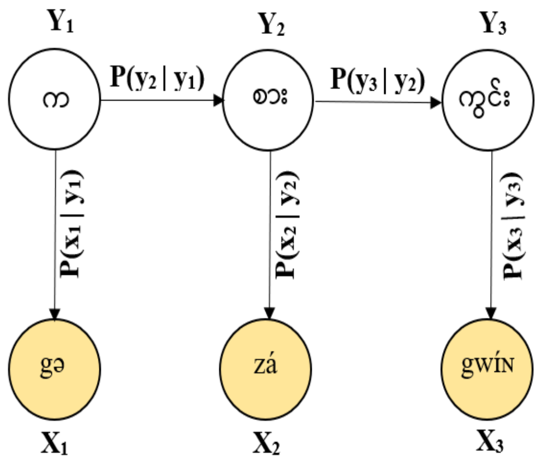

Fig. 2: An example of graphical representation of Hidden Markov Model

C. Conditional Random Field (CRF)

The Conditional Random Field (CRF) is a probabilistic model that directly models the posterior distribution of a label sequence conditioned on the observed data presented to it. CRF is similar to HMM, but CRF models do not make any assumptions about the independence or interdependence of the data being modeled. It uses attributes of the observed data to constrain the probabilities of the various labels that the observed data can receive.

The CRF defines a posterior probability $P(y|x)$ of a label sequence "y" for a given input sequence "x". The input sequence "x" corresponds to a series of syllable unit of Burmese text data, while the label sequence "y" is the series of IPA labels assigned to that observed syllable sequence. Each syllable in "x" is assigned exactly one IPA label in "y". The distribution of the IPA label sequence "y" given the observation sequence "x" will have the form in (2).

$$
P(Y|X) = \frac{1}{Z(x)} exp(\sum_{t=1}^{T}\sum_{k=1}^{|\lambda|} \lambda_k f_k(y_t, y_{t-1}, x_t)) \tag{2}
$$

In this equation, "t" ranges over the syllable indices of the observed data and Z(x) is a normalizing constant over all possible label sequences of "y" computed as in (3).

$$
Z(x) = \sum_{y} exp(\sum_{t=1}^{T}\sum_{k=1}^{|\lambda|} \lambda_k f_k(y_{t-1}, y_t, x_t)) \tag{3}
$$

The CRF is thus described by a set of feature functions (f_k), defined on graph cliques, with associated weights ($_k$) [21]. The example graphical representation is shown in Fig. 3.

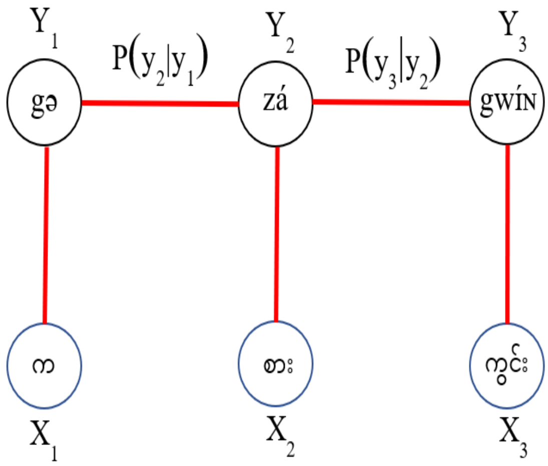

Fig. 3: An example of graphical representation of Conditional Random Field

D. Phrase-based Statistical Machine Translation (PBSMT)

A PBSMT translation model is based on joint phrasal units analogous to graphones. A length model, a language model on the target side and an re-ordering model are the components of a phrase-based translation system but the later model is not used for monotonic transduction such as G2IPA conversion. The models are integrated within a log-linear framework.

The phrase translation model based on the noisy channel model to find the best translation (e_{best}) that maximizes the translation probability $P(e|f)$ given the source sentences; mathematically. For example, the source language is French and the target language is English. The translation of a French sentence into an English sentence is modeled as (4) [22].

$$e_{best} = argmax_e P(e|f) \tag{4}$$

Applying the Bayes' rule, we can factorized the $P(e|f)$ into three parts as (5).

$$P(e|f) = \frac{P(e)}{P(f)} P(f|e) \tag{5}$$

The final mathematical formulation of phrase-based model is as (6).

$$argmax_e P(e|f) = argmax_e P(f|e)P(e) \tag{6}$$

E. Bi-directional Long Short-Term Memory (Bi-LSTM) Encoder-Decoder

The bi-directional recurrent neural network was proposed in [23]. In this architecture, one RNN processes the input from left-to-right, while another RNN processes it from right-to-left. The outputs of the two sub-networks are then combined. The idea has been used for speech recognition [23] and more recently for language understanding [24]. Bi-directional LSTMs have been applied to speech recognition [25] and machine translation [26]. In the bi-directional model, the phoneme prediction depends on the whole source-side letter sequence as in (7).

$$p(\phi_1^T|A, l_1^T) = \prod_{t=1}^{T} p(\phi_t|\phi_1^{t-1} l_1^T) \tag{7}$$

Figure 4 illustrates the example of this model. Focusing on the second set of inputs, for example, letter l_t = “စား” is projected to a hidden layer, together with the past phoneme prediction ϕ_{t-1} = gə. The letter l_t = “ စား” is also projected to a hidden layer in the network that runs in the backward direction. The hidden layer activation from the forward and backward networks is then used as the input to a final network running in the forward direction. The output of the topmost recurrent layer is used to predict the current phoneme ϕ_t = zá.

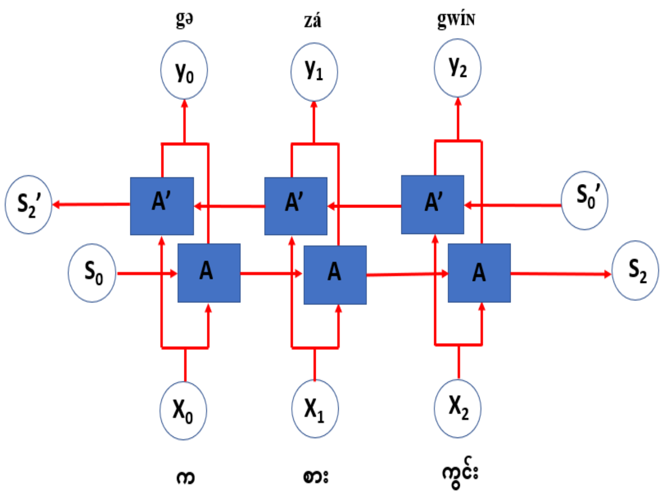

Fig. 4: Example of G2IPA Conversion of Myanmar word in Bi-directional Long Short-Term Memory(Bi-LSTM)

VII. Experimental Setup

A. Corpus Statistics

In the experiments, we used 24,803 G2IPA phoneme pairs. These pairs are developed based on myG2P (version 1.1). We randomized these pairs and split into 22,323 pairs for training, 2,480 pairs for open test sets for all RDR, HMM and CRF experiments. In PBSMT and Bi-LSTM (RNN) approaches, 19,843 pairs (80% of the corpus) were used for training, 2,480 pairs (10% of the corpus) for development and the last (10% of the corpus) 2,480 pairs for testing as shown in Table VIII. In this case, the combination of training and development (80% + 10% = 90%) used in the Bi-LSTM (RNN) approach equals to the training corpus of RDR, HMM and CRF approaches.

B. Frameworks used for Building RDR, HMM, CRF, PBSMT, and Bi-LSTM (RNN) models

We described the frameworks of RDR, HMM, CRF, PBSMT and Bi-LSTM (RNN) experiments for G2IPA conversion process.

TABLE VIII: Corpus Statistics of G2IPA Conversion Models

Model	**Data Size**
RDR	Train = 22,323 Test = 2,480
HMM	
CRF	
PBSMT	Train = 19,843 Development = 2,480 Test = 2,480
Bi-LSTM (RNN)	

1) RDRPOSTagger
RDRPOSTagger (version 1.2.4) is a rule-based Part-of-Speech and morphological tagging toolkit [27].
2) Trigram HMM
Jitar (version 0.3.3) is a simple part-of-speech tagger, based on a trigram Hidden Markov Model (HMM) [28], [29].
3) CRFSuite
We used the CRFSuite tool (version 0.12) [30] for training and testing G2IPA conversion CRF model. The main reason was that it can give the excellent training and testing speed than other CRF toolkits [31].
4) Moses Decoder
The Moses toolkit [32] is used for training the PBSMT statistical machine translation system. The syllable segmented grapheme/IPA symbols (source language) was aligned with the syllable segmented IPA symbols/grapheme (target language) using GIZA++ [33]. The alignment was symmetrized by grow-diag-final and heuristic [22]. The KenLM [34] language model was used for training the 5-gram language model with modified Kneser-Ney discounting [35]. Minimum error rate training (MERT) [36] was used to tune the decoder parameters and the decoding was perfromed using the Moses decoder (version 2.1.1) [32]. We used the default settings of decoder for PBSMT experiment.
5) Nagisa
We used Nagisa toolkit (version 0.2.7) [37], a python module for Japanese word segmentation/POS-tagging, to train Bi-LSTM network architecture for G2IPA conversion system. This tool based on Dynamic Neural Network (DyNet toolkit) to calculate the neural networks. And, the word segmentation model uses character- and word-level features [38]. However, in our conversion system, we used the segmentation unit as syllable and the IPA-tagging model as IPA tag dictionary information. We used default settings of Nagisa for Bi-LSTM training such as the dimensionality of hidden layer is set to 100 and RNN layer size is 1 layer. We trained Bi-LSTM model for maximum 10 epoches with the initial learning rate of 0.1 and the dropout rate of 0.3.

C. Evaluation

The performance of our proposed G2IPA conversion models are measured by accuracy, F1-score and PER.

1) Accuracy
Accuracy is that the most intuitive performance measure and it's simply a ratio of correctly predicted observation to the entire observations. The higher the accuracy, the better our model [39] and the accuracy can be calculated by using the following equation (8) [2].

$$Accuracy(\%) = (\frac{\#correct\ tags}{\#total\ tags}) * 100 \tag{8}$$

2) F1 Score
The **F-score** or **F-measure** is a measurement of the accuracy of the test set. It can be calculated from the precision and recall of the test set. F1 score is that the mean of precision and recall. It is calculated as shown in (9) [40].

$$F1 = (\frac{2 * Precision * Recall}{Precision + Recall}) \tag{9}$$

3) Phoneme Error Rate (PER)
For G2IPA phoneme conversion models, the converted outputs were analyzed using Phoneme Error Rate (PER) [41]. The SCLITE (score speech recognition system output) program from the NIST scoring toolkit SCTK (version 2.4.10) [42] was also used to make the dynamic programming-based alignments between reference (ref) and hypothesis (hyp) and calculation of WER. The WER calculation can be formulated as (10):

$$WER(\%) = (\frac{S + D + I}{S + D + C}) = (\frac{S + D + I}{N}) * 100 \tag{10}$$

where S is the number of substitutions, D is the number of deletions, I is the number of insertions, C is the number of correct words and N is the number of words in the reference ($N = S + D + C$) [41]. In our case, we trained our G2IPA conversion models with segmented syllable units and alignment process was done based on this units. Therefore, PER is derived from the above calculation at the phoneme level rather than the word level. The lower the PER, the better the conversion result is.

VIII. Results and Discussion

In Table IX and X, we reported G2IPA conversion performance based on the accuracies and F1 scores by comparing the experimental results of all models with myG2P (version 2.0) test data. The results shown in highlighted are the best scores among these five models.

Based on the experimental results, we discovered that the highest accuracies (98.70 and 97.70) and F1 scores (0.9869 and 0.9772) were achieved by CRF model for G2IPA conversion and reverse directions. In RDR and HMM conversion models, HMM model gave slightly higher accuracy (average +0.2) and F1 scores (average +0.0027) in G2IPA conversion than RDR as shown in Table VIII. For IPA2G conversion, HMM also got the better results (average +2.2 in accuracy and +0.0229 in F1 scores) as described in Table IX than RDR. In PBSMT and Bi-LSTM (one type of RNN) with 80% training data size of total corpus, PBSMT gave the stronger results with (average +3.64 and +5.72 accuracies) and (average +0.035 and +0.0596 F1 scores) than Bi-LSTM for both directions.

TABLE IX: Accuracies and F1 Scores of RDR, HMM, CRF, PBSMT, and Bi-LSTM (RNN) for G2IPA Conversion

Model	**Accuracy(%)**	**F1**
RDR	84.70	0.8467
HMM	84.90	0.8494
CRF	**98.70**	**0.9869**
PBSMT	86.80	0.8662
Bi-LSTM (RNN)	83.16	0.8312

TABLE X: Accuracies and F1 Scores of RDR, HMM, CRF, PBSMT, and Bi-LSTM (RNN) for IPA2G Conversion

Model	**Accuracy(%)**	**F1**
RDR	85.70	0.8543
HMM	87.90	0.8772
CRF	**97.70**	**0.9772**
PBSMT	88.50	0.8843
Bi-LSTM (RNN)	82.78	0.8247

There are two interesting points in the experimental results according to the comparison of G2IPA in Table VIII with IPA2G in Table IX. The first one is that the gains in accuracies and F1 scores of CRF and Bi-LSTM G2IPA conversion models decreased (average -1.0 and -0.38 in accuracy and average -0.0097 and -0.0065 in F1 scores) in the reverse conversion models. However, the accuracies and F1 scores of PBSMT, RDR and HMM conversion models in IPA2G conversion are higher (average +1.70, +1.0 and +3.0 in accuracy and +0.0181, +0.0076 and +0.0278 in F1 scores) than that of G2IPA ones. The second one is that PBSMT model got the second highest results among five models despite of the small amount of data used to train the conversion model.

From the overall results, CRF model is the best model for our myG2P (version 2.0) dictionary in both directions and PBSMT model is the second ones. The RDR, and HMM models have the comparable results and Bi-LSTM (RNN) model can be said as the lowest model in both conversions. This is because Bi-LSTM network architecture requires a lot of training data, and the model training process should also be tuned to various parameters.

IX. Error Analysis

We analyzed the converted outputs from all G2IPA conversion models using PER [35] with "SCLITE" command of SCTK toolkit [36]. The average PER results of all conversion models with the same test set for both directions are shown in Table XI.

The PER results show that G2IPA conversion of CRF and Bi-LSTM models gave the lower PER values than IPA2G conversion. The other three models achieved the lower PER values in IPA2G conversion.

TABLE XI: PER Results of G2IPA Conversion Methodologies

Model	**Grapheme-to-IPA PER(%)**	**IPA-to-Grapheme PER(%)**
RDR	15.30	14.30
HMM	15.10	12.10
CRF	**1.30**	**2.30**
PBSMT	13.20	11.50
Bi-LSTM (RNN)	16.84	17.22

The following PER calculations show the examples of some approaches on the converted outputs for G2IPA conversion with the same type of test set. The first one is an example PER calculation for CRF model. For example, the SCLITE program calculates the values of I, D, C, and S for Myanmar word "ခွန် စိုက် အားး စိုက်"("Work hard" in English)compare to a reference IPA symbols as follows:

Scores: (#C #S #D #I) 4 0 0 0
REF: kʰʊ̀ɴ zaɪʔ ʔá zaɪʔ
HYP: kʰʊ̀ɴ zaɪʔ ʔá zaɪʔ
Eval:

In this case, there are no insertion, deletion and substitution and all converted outputs are correct. Therefore, PER value is 0.

The next example is for Bi-LSTM (RNN) model and there is only one correct symbol and the values of substitution are 3 (Kʰʊ̀ɴ ==> Tɕḭ, and two substitutions of ZAɪʔ ==> KAʊʔ). Thus, its PER value is 75%.

Scores: (#C #S #D #I) 1 3 0 0
REF: Kʰʊ̀ɴ ZAɪʔ ʔá ZAɪʔ
HYP: Tɕḭ KAʊʔ ʔá KAʊʔ
Eval: S S S

After we had made the detail analysis of the confusion pairs for all conversion models, we found that most of the confusion pairs are occurred by (1) the tone errors (creaky, high and low tone), (2) consonant errors (also known as context dependent errors) and the vowel combination errors. For example, the top 10 confusion pairs of G2IPA conversion based RDR, HMM, CRF, PBSMT, and BI-LSTM (RNN) models are shown in Table XII (a), (b), (c), (d), and (e).

In these sub-tables, the 1^{st} and 2^{nd} columns are the reference-hypothesis pair and the value of error frequent (i.e. output of each conversion model) for G2IPA conversion. In RDR model, most of the confusion pairs are caused by the unchangeable syllable pronunciation with neighbouring context (also known as consonant errors), for example, "də ==> tə" and "paʊʔ ==> baʊʔ". In HMM model, there are also consonant errors such as "də ==> tə" and "paʊʔ ==> baʊʔ". There is another type of error such as "jè ==> jì". This is called the vowel combination error and caused by the pronunciation nature of Myanmar people.

TABLE XII: Top Ten Confusion Pairs of G2IPA Conversion Methodologies

(a) RDR

Ref ==> Hyp	Freq
də ==> tə	14
paʊʔ ==> baʊʔ	14
gə ==> kə	13
ká ==> gá	10
sə ==> zə	10
tɕí ==> dʑí	10
gà ==> kà	9
zaʔ ==> saʔ	9
zìɴ ==> sìɴ	9
bə ==> pə	8

(b) HMM

Ref ==> Hyp	Freq
paʊʔ ==> baʊʔ	13
zà ==> sà	11
də ==> tə	10
jè ==> jì	8
gə ==> kə	7
tə ==> də	7
zaʔ ==> saʔ	7
dʑa̰ ==> tɕa̰	6
gə ==> gá	6
ká ==> gá	6

(c) CRF

Ref ==> Hyp	Freq
wú ==> wó	2
ba ==> pə	1
bjḭ ==> bɛ̰	1
bjɔ̀ ==> dwè	1
da ==> tàɴ	1
daíɴ ==> dáɪɴ	1
do̰ʊɴ ==> tʰɛ́	1
dwé ==> θwé	1
dəɹà ==> sʰó	1
dəɹə ==> ŋə	1

(d) PBSMT

Ref ==> Hyp	Freq
paʊʔ ==> baʊʔ	13
də ==> tə	11
zaʔ ==> saʔ	8
jè ==> jì	7
tɕí ==> dʑí	7
gáɴ ==> kʰáɴ	6
gə ==> gá	6
gə ==> kə	6
ká ==> gá	6
gàɴ ==> kʰàɴ	5

(e) Bi-LSTM (RNN)

Ref ==> Hyp	Freq
də ==> tə	13
gə ==> kə	10
tə ==> də	10
bà ==> pà	7
gàɴ ==> kʰàɴ	7
ja̰ ==> jə	7
ma̰ ==> mə	7
sà ==> zà	7
tɕí ==> dʑí	7
zí ==> sí	7

In CRF conversion model, the confusion paris of "wú ==> wó" and "dəɹə ==> ŋə" are caused by the consonant errors and that of "bjḭ ==> bɛ̰" and "dwé ==> θwé" are due to vowel combination errors. In PBSMT model, the most common confusion pairs are occurred by the consonant error, for example, "paʊʔ ==> baʊʔ", "də ==> tə" and so on. Some pairs are also due to the pronunciation nature of Myanmar people such as "jè ==> jì".

In Bi-LSTM (RNN) model, most of the confusion pairs are happened by the consonant errors, for example, "də ==> tə", "zí ==> sí". Other confusion pairs of "ja̰ ==> jə" and "ma̰ ==> mə" are caused by tone errors (creaky tone and tone in minor syllable).

X. Conclusion

In this paper, the results of G2IPA and the reverse directions are presented by applying RDR, HMM, CRF, PBSMT, and Bi-LSTM (RNN) approaches, respectively. For the data, we manually labeled IPA symbols for all Burmese words of myG2P dictionary and used them for all conversion experiments. This dictionary contains over 24K G2IPA phoneme pairs and released as myG2P (version 2.0) for Burmese ASR and TTS R&D developments (https://github.com/ye-kyaw-thu/myG2P/tree/master/ver2). From the overall experimental results, CRF model achieved the highest accuracy, F1 scores, and the lowest PER values for both conversions. The PBSMT model got the second highest results. The RDR and HMM models had the comparable results and Bi-LSTM model gained the lowest scores in both directions. The calculation of PER values and some confusion pairs are also explained in details and the most common pairs are caused by consonant and vowel combination errors. In the near future, we plan to check and modify our G2IPA phoneme pairs by adding the new words from other domains to solve the confusion pairs and types of errors. We will apply myG2P (version 2.0) dictionary in Burmese ASR and TTS research works to prove that this dictionary can be very helpful in these areas. We will also extend our experiments with other sequence learning approaches such as Neural Conditional Random Field (NCRF).

Acknowledgment

We would like to special thanks to the first contributors of myG2P (version 1.1) for getting a chance of extending as "Grapheme-to-IPA" phoneme pairs dictionary myG2P (version 2.0). We would like to thanks U Nyi Hla Nge foundation, Yangon Technological University, Yangon, Myanmar for their supporting of research funds to our research work. We would also like to thank all participants for their kind contributions to our research.

References

[1] Ei Phyu Phyu Soe, "Grapheme-to-Phoneme Conversion for Myanmar Language," In Proceedings of the 11th International Conference on Computer Applications (ICCA), Yangon, Myanmar, February, 2013, pp. 195-200.

[2] Ye Kyaw Thu, Win Pa Pa, Finch Andrew, Aye Mya Hlaing, Hay Mar Soe Naing, Sumita Eiichiro, and Hori 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.

[3] Ye Kyaw Thu, Win Pa Pa, Finch Andrew, Ni Jinfu, Sumita Eiichiro, and Hori Chiori, "The Application of Phrase based Statistical Machine Translation Techniques to Myanmar Grapheme to Phoneme Conversion," in the Pacific Association for Computational Linguistics Conference (PACLING), Legian, Bali, Indonesia, May, 2016, pp. 170-176.

[4] Ye Kyaw Thu, Win Pa Pa, Yoshinori Sagisaka, and Naoto 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.
[5] Kanishka Rao, Fuchun Peng, Hasim Sak, and Framcoise Beaufays, "Grapheme-To-Phoneme Conversion Using Long Short-Term Memory Recurrent Neural Networks," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, QLD, Australia, April, 2015.
[6] http://en.wikipedia.org/wiki/Burmese_language
[7] http://r12a.github.io/scripts/myanmar/
[8] Myanmar Sar A Phwae, Myanmar Spelling Bible, "Myanmar Thut Pon Kyan", Yangon, 2003."
[9] Myanmar Language Committee, "Myanmar Grammar," Ministry of Education, Myanmar, 2005.
[10] Thein Tun, "Some acoustic properties of tones in Burmese" In D. Bradley (ed), Papers in South- East Asian Linguistics8: Tonation Canberra: Australian National University, 1982, pp. 77-116.
[11] https://en.wikipedia.org/wiki/Burmese_alphabet
[12] Department of the Myanmar Language Commission, Myanmar-English Dictionary, Ministry of Education, Yangon, 1993.
[13] http://github.com/ye-kyaw-thu/sylbreak
[14] https://tech.labs.oliverwyman.com/blog/2019/11/22/ripple-down-rules/
[15] Compton, P., and Jansen, R., "A philosophical basis for knowledge acquisition," Knowledge acquisition, 2(3), 1990, pp. 241-258.
[16] Compton, P., and Jansen, R., "A philosophical basis for knowledge acquisition," Knowledge acquisition, 2(3), 1990, pp. 241-258.
[17] Scheffer and Tobias, "Algebraic Foundation and Improved Methods of Induction of Ripple Down Rules," 1996, pp. 23-25.
[18] Durbin, R., Eddy, S., Krogh, A., and Mitchinson, G., "Biological sequence analysis: probabilistic models of proteins and nucleic acids," Cambridge University Press, Cambridge, 1998.
[19] Rabiner, L.R., "A tutorial on hidden Markov models and selected applications in speech recognition," in Proceedings of the IEEE, 77(2), 1989, pp. 257-286.
[20] Jurafsky, Daniel, and James H.Martin, "Speech and Language Processing: An Introduction to Natural Language Processing," Computational Linguistics, and Speech Recognition, 1st edition, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2000.
[21] J. J. Morris and E. F. Lussier, "A Study on the use of Conditional Random Fields for Automatic Speech Recognition," Department of Computer and Information Science, Ohio State University, Neil Avenue Columbus, OH, United States, 2010.
[22] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation." In Proceedings of HTL-NAACL, 2003, pp. 48–54.
[23] M. Schuster and K. Paliwal, "Bidirectional recurrent neural networks," IEEE Trans. on Signal Processing, vol. 45, no. 11, 1997, pp. 2673-2681.
[24] G. Mesnil, X. He, L. Deng, and Y. Bengio, "Investigation of recurrent-neural-network architectures and learning methods for language understanding," in INTERSPEECH, 2013.
[25] A. Graves, "Generating sequences with recurrent neural networks," 2013.
[26] M. Sundermeyer, T. Alkhouli, J. Wuebker, and H. Ney, "Translation modeling with bidirectional recurrent neural networks," in Emprical Methods in Natural Language Processing (EMNLP), 2014.
[27] D. Q. Nguyen, D. Q. Nguyen, D. D. Pham, and S. B. Pham. "RDRPOSTagger: A Ripple Down Rules-based Part-Of-Speech Tagger," In Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics, 2014, pp. 17-20.
[28] Brants and Thorsten, "TnT A Statistical Part-of-speech Tagger," In Proceeding of the 6th Conference on Applied Natural Language Processing, Stroudsburg, PA, USA. Association for Computational Linguistics (ACLC), 2000, pp. 224-231.
[29] De Kok, Daniël 2014. "Jitar: A simple Trigram HMM part- of-speech tagger," [accessed 2016].
[30] Noaki Okazaki, "Crfsuite: a fast implementation of conditional random fields (crfs)," 2007.
[31] https://github.com/chokkan/crfsuite
[32] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst, Moses: Open Source Toolkit for Statistical Machine Translation, Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session, Prague, Czech Republic, June, 2007.
[33] O. F. Josef and N. Hermann, "Improved Statistical Alignment Models", In Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, Hong Kong, China, 2000, pp. 440-447.
[34] Heafield, Kenneth, "KenLM: Faster and Smaller Language Model Queries", In Proceedings of the Sixth Workshop on Statistical Machine Translation, WMT 11, Edinburgh, Scotland, 2011, pp. 187-197.
[35] C. Stanley and G. Joshua, "An empirical study of smoothing techniques for language modeling", In Proceedings of the 34th annual meeting on Association for Computational Linguistics, 1996, pp. 310-318.
[36] O. Franz, "Minimum error rate training in statistical machine translation", In Proceedings of the 41st Annual Meeting and Association for Computational Linguistics Volume 1, Sapporo, Japan, July, 2003, pp.160-167.
[37] https://github.com/taishi-i/nagisa
[38] https://nagisa.readthedocs.io/en/latest/index.html
[39] https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/
[40] https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/
[41] Wikipedia of Word Error Rate: https://en.wikipedia.org/wiki/Word_error_rate
[42] (NIST) The National Institute of Standards and Technology. Speech recognition scoring toolkit (sctk), version: 2.4.10, 2015.

Honey Htun received the B.E. degree in Information Technology at Technological University (Kyaukse) in 2005, and M.E degree in Information Technology from Mandalay Technological University (MTU) in 2015. In 2009, she joined the Department of Information Technology, Technological University (Kyaukse), as a Demonstrator, and in 2017 she became an Assistant Lecturer at the Department of Computer Engineering and Information Technology, Yangon Technological University. She is also a Ph.D. candidate in Computer Engineering and Information Technology (CEIT) Department of Yangon Technological University (YTU). She focused her efforts on her doctoral thesis research of Myanmar Speech to English Text Translation System for Primary Education.

Ni Htwe Aung received the B.E. degree in Information Technology at Technological University (Monywa) in 2007, and M.E degrees in Information Technology from Technological University (Monywa) in 2010. In 2011, she joined the Department of Information Technology, Technological University (Monywa), as an Instructor, and in 2018 she became an Assistant Lecturer at the Department of Computer Engineering and Information Technology, Yangon Technological University. She is also a Ph.D candidate in Computer Engineering and Information Technology (CEIT) Department of Yangon Technological University (YTU). She focused her efforts on her doctoral thesis research of Real-time Myanmar Fingerspelling Recognition.

Shwe Sin Moe received the B.E. degree in Computer Engineering at University of Technology (Yatanarpon Cyber City) in 2016. In 2018, she joined the Department of Computer Engineering and Information Technology, Yangon Technological University, as a Demonstrator. She is also a Master candidate in Computer Engineering and Information Technology (CEIT) Department of Yangon Technological University (YTU). She focused her efforts on her Master thesis research of Text-to-Speech (TTS) Conversion System for Romanized Myanmar Text.

Wint Theingi Zaw received the B.E. degree in Information Technology at Technological University (Pathein) in 2008. In 2009, she joined the Department of Information Technology, Technological University (Taungoo), as a Demonstrator, in 2014, she transferred into the Department of Information Technology, Technological University (Pathein), and in 2017 she became an Assistant Lecturer at the Department of Computer Engineering and Information Technology, Yangon Technological University. She is also a Master candidate in Computer Engineering and Information Technology (CEIT) Department of Yangon Technological University (YTU). She focused her efforts on her Master thesis research of Machine Translation Based Myanmar Romanization.

Nyein Nyein Oo is a Professor at the Department of Computer Engineering and Information Technology, Yangon Technological University, Yangon, Myanmar. She is currently working as a Head of Department and a leader of Data Science research team. Her current responsibilities are managing the department, supervision and teaching the undergraduate and postgraduate students. She earned her Ph.D. in Information Technology from the Yangon Technological University in 2006. Her research interest is focused on parallel and distributed computing and data science especially in Natural Language Processing (NLP).

Thepchai Supnithi received the B.S. degree in Mathematics from Chulalongkorn University in 1992. He received the M.S. and Ph.D. degrees in Engineering from the Osaka University in 1997 and 2001, respectively. He is currently head of language and semantic research team artificial intelligence research unit, NECTEC, Thailand.

Ye Kyaw Thu is a Visiting Professor of Language & Semantic Technology Research Team (LST), Artificial Intelligence Research Unit (AINRU), National Electronic & Computer Technology Center (NECTEC), Thailand and Head of NLP Research Lab., University of Technology Yatanarpon Cyber City (UTYCC), Pyin Oo Lwin, Myanmar. He is also a founder of Language Understanding Lab., Myanmar and a Visiting Researcher of Language and Speech Science Research Lab., Waseda University, Japan. He is actively co-supervising/supervising undergrad, masters' and doctoral students of several universities including KMITL, SIIT, UCSM, UCSY, UTYCC and YTU.