Grapheme to Syllable Sequence Phoneme Conversion for Burmese Spelling TTS

Hnin Yu Hlaing, Ye Kyaw Thu, Hlaing Myat Nwe, Hnin Aye Thant,
Htet Ne Oo, Thepchai Supnithi, Aung Win

***Abstract*— Grapheme-to-Phoneme (G2P) conversion is the most important stage for building automatic speech recognition (ASR) and text-to-speech (TTS) systems. This work is an essential part of the Burmese (Myanmar language) spelling TTS project. This paper addresses the problem of grapheme to phoneme conversion for spelling Burmese primer has not yet been investigated. Burmese words are formed by one or more than one syllable and a syllable is composed of consonant, consonant cluster, vowel and vowel diacritics. According to the nature of Burmese primer, a syllable is spelled out in order of its structure. Therefore, in our experiment, we have to use two main parts of conversion: syllable to spelling and spelling to phoneme. We propose three grapheme to phoneme conversion methods: Averaged Perceptron, Ripple Down Rules (RDR), Conditional Random Field (CRF). The experimental results of these three approaches were evaluated with phoneme error rate (PER), Bilingual Evaluation Understudy (BLEU) and Character n-gram F-score ($chrF^{++}$). The result shows that RDR is the best performing approach with the lowest error rate (1.0% and 0.1%) for both conversions on reading Burmese primer. CRF achieved 100% BLEU and $chrF^{++}$ scores on a closed test of spelling to phoneme conversion. The open-test evaluation scores of Averaged Perceptron and CRF also achieved over 95% in syllable to spelling conversion and 99% in spelling to phoneme conversion.**

***Index Terms*—Grapheme-to-Phoneme, G2P, Averaged Perceptron, RDR, CRF, Burmese (Myanmar language)**

I. Introduction

GRAPHEME-to-Phoneme (G2P) conversion method is needed to develop natural language processing, text-to-speech synthesis (TTS), and automatic speech recognition (ASR) systems and has been widely researched with a variety of different techniques. G2P is the process of generating phonetic transcription from the written form of text. The one-to-one correspondence between written text and spoken sound is the simplest way. However, the same written text has different pronunciation depending on the surrounding context. It is difficult to translate a word to its pronunciation for such a case. Burmese is an low-resourced language and a syllable has different pronunciations depending on the context and its Part-of-Speech (POS). Most machine learning approaches for G2P conversion are applied with supervised learning methods. In this paper, we experiment with three G2P conversion methods on spelling Burmese primer. We intended this experiment for Kindergarten and Primary level students and students at the non-formal educational platform. Our experiment focuses on the spelling style of Burmese primer and therefore, we collected short sentences of Burmese Textbooks from Kindergarten to Grade-8 including Paliloan words. Then we used automatic evaluation criteria of PER for the performance of G2P conversion.

An overview of the contents of this paper is as follows. Data preparation for our experiment was manually collected based on the spelling style of Burmese primer (Section II-A). Methodologies used in the experiment are described in Section IV. Data collection, preparation, software used, and evaluation detailed are explained in Section V, and result and errors getting from the methods are analyzed in Section VI and VII. The conclusion is described Section VIII.

II. Introduction to Burmese

Burmese is one of Lolo-Burmese grouping of the Sino-Tibetan language family spoken by 32 million people in Myanmar where it is an official language and the language of Myanmar people. Like all Sino-Tibetan languages, Burmese has a simple syllable structure consisting of an initial consonant or consonant cluster followed by a vowel with an associated tone. It is a tonal, pitch-register and monosyllabic language (high, low and creaky) and two other tones (stopped and reduced). In Burmese text, one syllable is composed of consonant, consonant cluster, vowel and vowel diacritics in the form of C(G)V((V)C) structure. For example,

1) CV - မယ် (“young women” in English)
2) CVC - မက် (“crave” in English)
3) CGV - ငြိ(“earth” in English)
4) CGVC - မျက် (“eye” in English)
5) CVVC - မောင် (“term of address young men” in English)

Hnin Yu Hlaing is with the NLP Lab., University of Technology (Yadanarpon Cyber City)), Myanmar.

Ye Kyaw Thu and Thepchai Supnithi are with National Electronics and Computer Technology Center, Thailand.

Hlaing Myat Nwew, Hnin Aye Thant, Aung Win and Htet Ne Oo are with University of Technology (Yadanarpon Cyber City), Myanmar.

Corresponding Authors: hninyuhlaing@utycc.edu.mm and yktnlp@gmail.com

Manuscript received February 8, 2021; accepted March 6, 2021; revised March 13, 2021; published online April 1, 2021.

6) CGVVC - မြောင်း (“ditch” in English)

TABLE I: Myanmar Scripts and Consonants

Consonants			
Unaspirated	**Aspirated**	**Voiced**	**Nasal**
က /k/	ခ /kh/	ဂ /g/, ဃ /g/	င/ng/
စ /s/	ဆ /hs/	ဇ /z/, ၐ /z/	ည/nj/
ဋ /t/	ဌ /ht/	ဍ /d/, ဎ /d/	ဏ/n/
တ /t/	ထ /ht/	ဒ /d/, ဓ /d/	န/n/
ပ /p/	ဖ /hp/	ဗ /b/, ဘ /b/	မ/m/
ယ /j/	ရ /j/	လ /l/, ဝ /w/	သ/th/
	ဟ /h/	ဠ /l/, အ /a/	

Table I shows the group of characters according to their sound of letter and arrange in the traditional order. And the following are 12 vowels, 4 medials, and independent vowels.

Vowels

အ (a.), အိ (i.), အီ (i), အု (ou.), အူ (ou), ဧ (ae), အဲ (e:), ဪ (au.), ဩ (au), အံ (an), အား (a:), အက် (e’)

Medials

- ယပင့် (ja. pin.) - Written ျ
- ယရစ် (ja. ji’) - Written ြ
- ဝဆွဲ (wa. hswe:) - Written ွ
- ဟထိုး (ha. htou:) - Written ှ

Independents Vowels

ဣ (i.), ဤ (i), ဥ (ou.), ဦ (ou), ဧ (a), ဩ (au:), ဪ (au)

A. Spelling Style of Burmese Primer Reader

The most basic teaching of Burmese language is Burmese primer called Thin Poun Gyi that must be taught in Grade-1. The Burmese alphabet consists of 33 consonants, 4 medials, 12 vowels, other symbols, and special characters, and is written from left to right. In the spelling system of Burmese primer, a syllable is spelled out each character name based on the syllable formation structure and read the pronunciation of the syllable at the end. An example of spelling system of Burmese primer, ကို (kou) is spelled out ကကြီး (က - consonant), လုံးကြီးတင် (ိ - diacritic symbol), တစ်ချောင်းငင် (ု diacritic symbol) and ကို in order of syllable structure. Therefore, grapheme to phoneme conversion of spelling Burmese primer system is developed with two parts in this paper. The first part is syllable to spelling conversion (ကို to ကကြီး-လုံးကြီးတင်-တစ်ချောင်းငင်-ကို) and the second is spelling to phoneme conversion (ကကြီး-လုံးကြီးတင်-တစ်ချောင်းငင်-ကို to ka. gyi: loun: gyi: tin ta- chaun: ngin kou).

III. Related Work

There has been extensive research on grapheme to phoneme conversion from rule-based approaches to the current state-of-the-art deep learning based approach. The dictionary-based Burmese G2P conversion approach is firstly proposed by [1] and this approach worked on only Burmese syllables with only 133 phonemes to speech out for all Burmese texts. They proposed dictionary based approach and rule based approach based on Burmese phonological rules. They reported the evaluation of the quality of G2P conversion with PER with resulting accuracy of 93.54 according to their phonological rules. In the study of [2], the authors compared the methods of G2P conversions using different machine learning models for the Burmese. They experimented and evaluated on seven different methods: Adaptive Regularization of Weight Vectors (AROW) based structured learning (S-AROW), CRF, Joint-sequence models (JSM), phrase-based statistical machine translation (PBSMT), Recurrent Neural Network (RNN), Support Vector Machine (SVM) based pointwise classification and Weighted Finite State Transducers (WFST) on pronunciation dictionary. Their results proposed that CRF, PBSMT and WFST approaches are the best methods of G2P conversion on Burmese according to PER and manual checking. Ye Kyaw Thu et al. [3] expressed CRF approach in G2P conversion based on syllable pronunciation features with eight patterns. They experimented the largest improvement in the word and phoneme accuracy by adding the combinations of features to the baseline model. In [4], different stochastic tagging schemes were proposed for tagging efficiency on part of speech tagger. They suggested that the perceptron tagger provided the best suitable tagger in terms of both evaluation scores and run time. RDR and WFST approaches were applied by [5] for Burmese name romanization. In their experiment, RDR provided slightly better accuracy than with open test data set and WFST achieved better accuracy on the closed test data set. Khaing Hsu Wai et al., solved String Similarity measures based on phoneme similarity for Burmese strings in [6]. The authors reported that there was a better word correction error rate of their string similarity measurements on all existing distance measures based on G2P mapping. Denis Jouvet et al. presented G2P conversion with CRF and Joint-Multigram Model (JMM) for automatic speech recognition. They described that CRF-based approach lead to better result of pronunciation error rate while generating pronunciation variant per word and the two combinations gave to improved speech recognition performance rather than the use of single approach in multiple pronunciation variants [7]. Vathnak Sar et al. examined a simple way of adding linguistic knowledge into the statistical G2P convertor by simply inserting three types of vowel tags into a Khmer word [18].

IV. Methodology

In this section, we express the G2P conversion methodologies used for the experiment in this paper.

A. Averaged Perceptron

Averaged perceptron is a modification of perceptron algorithm for improving training accuracy or convergence and overfitting problem of perceptron [19], [27]. It is supervised learning method. It is more stable and generates accurate result relative to the perceptron algorithm. In this scheme, linear function classifies inputs

into several possible outputs, and then a set of weights that are obtained from feature vector throughout all iteration is averaged. In each iteration, weight coefficients are updated. Weight coefficients for all features based on a current sentence and algorithm tags are decreased by 1. Weight coefficients for all features based on a current sentence and correct tags are increased by 1 [20]. So, if algorithm tags are all correct, weight coefficients remain unchanged. The task of the Averaged Perceptron algorithm is to sum up all the coefficients of true features in a given context [22]. This equation (1) can be expressed as

$$\omega(C,T) = \sum_{i=1}^{n} \alpha_i . \phi_i(C,T) \quad (1)$$

where $\omega(C,T)$ is the transition weight for tag T in context C, n is the number of features, α_i is the weight coefficient of the i^{th} feature and ϕ_i (C, T) is the evaluation of the i^{th} feature for context C and tag T [22]. We prepared the annotated grapheme and phoneme data to model the features using a feature functions which estimate the featured weight of grapheme and phoneme pair during the training phase. The basic feature set of the Averaged Perceptron is

$$\phi(h_i, t_i) = \begin{cases} 1 & \text{if } t_i = \text{ka. and } w_i = \text{က} \text{ and } i \neq n \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where t is the current phoneme ka., (w_i) is the current word က, and the current word is not the last in the sentence. h is the history of feature functions of the previous two tags.

B. Ripple Down Rules (RDR)

RDR makes knowledge acquisition (KA) to add incremental knowledge to the system's knowledge without degrading the previous knowledge [25], [26]. Single Classification Ripple Down Rules (SCRDR) is a binary tree with two different kinds of edges : except and if-not edges. Each node in a tree is associated with a rule. The form of a rule is: if A then B where A and B are called the condition and the conclusion. Nested if statements are contained as except conditions shown in Figure 1.

The initial node is default rule or root without conditions. This rule is true so that it can link to the second node. If this second node's condition is not met, alternate if-not link to the another node. The intended conclusion is finally reached with if-then-exception statements. Some of the examples of G2P conversion for syllable to spelling can be seen as follows.

- The two syllables ယဉ် and ကျေး ("polite" in English) are spelled ယပက်လက်-ဥလေးသတ်-ယဉ် and သဝေထိုး-ကကြီး-ရပင့်-ဝစ်စနှစ်လုံးပေါက်-ကျေး. Therefore, we have two spelling sequences conclusions: ယပက်လက်-ဥလေးသတ်-ယဉ် and သဝေထိုး-ကကြီး-ရပင့်-ဝစ်စနှစ်လုံးပေါက်-ကျေး.

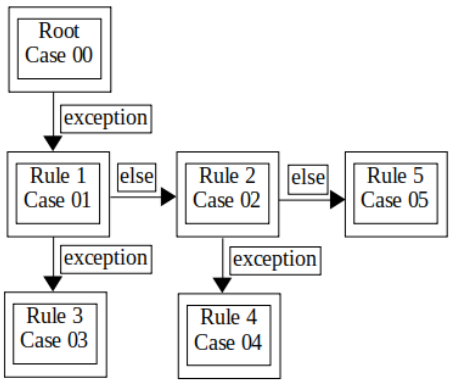

Fig. 1: A Binary Tree Classification of Single Classification of Ripple Down Rules

- The two syllables ပု and ဆိုး ("cloth worn by males" in English) are concluded as the two spelling sequences: ပစောက်-တစ်ချောင်းငင်-ပု and ဆလိမ်-လုံးကြီးတင်-တစ်ချောင်းငင်-ဝစ်စနှစ်လုံးပေါက်-ဆိုး.
- The conclusions of the two syllables ညီ and လေး ("little brother" in English) are ည-လုံးကြီးတင်-ဆန်ခတ်-ညီ and သဝေထိုး-လ-ဝစ်စနှစ်လုံးပေါက်-လေး.

Next, the spelling sequences to phoneme conversion is:

- ယ ပက် လက် ဥ လေး သတ် ယဉ် သ ဝေ ထိုး က ကြီး ရ ပင့် ဝစ် စ နှစ် လုံး ပေါက် ကျေး > ja. pe' le' nja. lei: tha' jin tha- wei htou: ka. gyi: ja. pin. wi' sa. nha- loun: pau' kyei:
- ပ စောက် တစ် ချောင်း ငင် ပု ဆ လိမ် လုံး ကြီး တင် တစ် ချောင်း ငင် ဝစ် စ နှစ် လုံး ပေါက် ဆိုး > pa. zau' ta- chaun: ngin pu. hsa. lein loun: gyi: tin ta- chaun: ngin wi' sa. nha- loun: pau' hsou:
- ည လုံး ကြီး တင် ဆန် ခတ် ညီ သ ဝေ ထိုး လ ဝစ် စ နှစ် လုံး ပေါက် လေး > nja. loun: gyi: tin hsan kha' nji tha- wei htou: la. wi' sa. nha- loun: pau' lei:

C. Conditional Random Field (CRF)

Conditional Random Field (CRF) is a classifier to predict the contextual information or state of neighbors affect the current prediction [28]. In this experiment, features are extracted from the attributes in a data set. In this experiment, the feature set of attributes used in CRF is $\{w_{t-2}, w_{t-1}, w_t, w_{t+1}, w_{t+2}\}$(where t is the index of the syllable being labeled). For example of syllable to spelling conversion, the feature sets of the syllable လည်း in the sentence of ငါးဖယ်လည်း ပါ၏ (It also contains featherback) is

လ-ညသတ်-ဝစ်စနှစ်လုံးပေါက်-လည်း w[-2] = ငါး w[-1]=ဖယ် w[0]=လည်း w[1]w[2]= w[-1]|w[0]=ဖယ်|လည်း w[0]|w[1]=လည်း|ပါ pos[-2]=c-ရေးချ-ဝစ်စနှစ်လုံးပေါက်-ငါး pos[-1]=ဖဦးထုပ်-ယပက်လက်သတ်-ဖယ် pos[0]=လ-ညသတ်-ဝစ်စနှစ်လုံးပေါက်-လည်း pos[1]=ပစောက်-ရေးချ-ပါ pos[2]=အက်ခရာ-၏ pos[-2]|pos[-1]=c-ရေးချ-ဝစ်စနှစ်လုံးပေါက်-ငါး|ဖဦးထုပ်-ယပက်လက်သတ်-ဖယ် pos[-1]|pos[0]=ဖဦးထုပ်-ယပက်လက်သတ်-ဖယ်|လ-ညသတ်-ဝစ်စနှစ်လုံးပေါက်-လည်း pos[0]|pos[1]=လ-ညသတ်-ဝစ်စနှစ်လုံးပေါက်-လည်း|ပစောက်-ရေးချ-ပါ pos[1]|pos[2]=ပစောက်-ရေးချ-ပါ|အက်ခရာ-၏ pos[-2]|pos[-1]|pos[0]=c-ရေးချ-

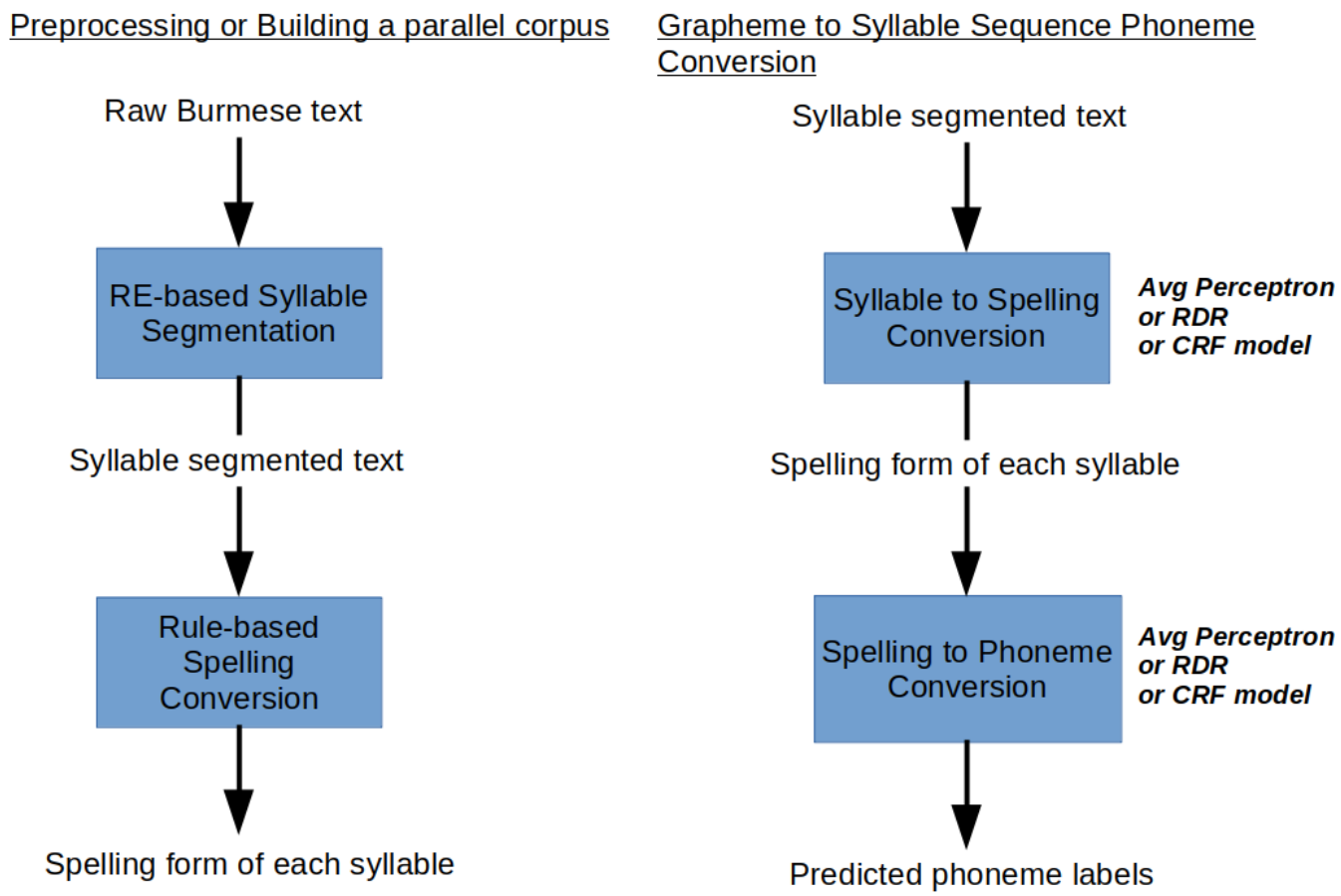

Fig. 2: The Overview of the Grapheme to Syllable Sequence Phoneme Conversion for Burmese spelling

ဝစ်စနှစ်လုံးပေါက်-ငါး|ဖဦးထုပ်-ယပက်လက်သတ်-ဖယ် |လ-ညသတ်-ဝစ်စနှစ်လုံးပေါက်-လည်း pos[-1]|pos[0]|pos[1]=ဖဦးထုပ်-ယပက်လက်သတ်-ဖယ်|လ-ညသတ်-ဝစ်စနှစ်လုံးပေါက်-လည်း|ပစောက်-ရေးချ-ပါ pos[0]|pos[1]|pos[2]=လ-ညသတ်-ဝစ်စနှစ်လုံးပေါက်-လည်း|ပစောက်-ရေးချ-ပါ|အက်ခရာ-၏

These types of feature sets are also applied in spelling to phoneme conversion.

D. Overall Structure of Grapheme to Syllable Sequence Phoneme Conversion

Figure 2 illustrates the overall structure of grapheme to syllable sequence phoneme conversion of Burmese Spelling TTS.

As the preprocessing tasks, input Burmese texts are segmented into syllables. And these segmented syllables are spelled with Perl script. For example,

Input Text: ကလေးများ (Children)
Syllable Segmented: က လေး များ
Spelling Converted: ကကြီး-က သဝေထိုး-လ-ဝစ်စနှစ်လုံးပေါက်-လေး မ-ရပင့်-ရေးချ-ဝစ်စနှစ်လုံးပေါက်-များ

For numbers,

Input Burmese Text: ၁၀ (Ten)
Syllable Segmented: ၁၀
Spelling Converted: တစ်ဆယ်

For Pali-loan word,
Input Text: စက္ကူစ (A piece of paper)
Syllable Segmented: စက္ကူ စ
Spelling Converted: စလုံး-ကသတ်-စက်-ကကြီး-နှစ်ချောင်းငင်-ကူ စလုံး-စ

The syllable to spelling conversion step is the very first and important for the next step spelling to phoneme conversion. In this step, a syllable is mapped to its spelling sequence. The longest spelling sequence of one syllable has five components of C (consonant), G (consonant cluster) and V (vowel). Example sequence of the syllable "ကျိုး" ("broken" in English) that contains "ကျ" (CG) and "ိ ု း" (V), is spell as ကကြီး-ရပင့်-လုံးကြီးတင်-တစ်ချောင်းငင်-ဝစ်စနှစ်လုံးပေါက်-ကျိုး. The spelling of the syllable is very simple. It does not depend on the context of the syllable. We gained many other syllables when a syllable is spelled. Therefore, we chose short sentences for our experiment. In the step of spelling to phoneme conversion, spelling sequences from the syllable to spelling conversion are

segmented into syllables and these syllables are mapped to its phonemes. We manually prepared some training data where the pronunciation of a syllable has different phonemes. Example of spelling to phoneme conversion is as follows:

Input Text: သဝဲထိုး-နဃ်-နေ မ-ငသတ်-ဝစ်စနှစ်လုံးပေါက်-မင်း (Sun)
Syllable Segmented: သ ဝဲ ထိုး နဃ် နေ မ ငသတ် ဝစ် စ နှစ် လုံး ပေါက် မင်း
Phoneme Labels: tha- wei htou: na. nge nei ma. nga. tha' wi' sa. nha- loun: pau' min:

Averaged Perceptron, RDR and CRF methods are applied to predict phoneme labels on both conversions.

V. Experimental Setup

A. Data Preparation

In the experiment, we collected 7,000 sentences in Burmese Textbooks from Kindergarten to Grade-8. It contains 1,639 unique syllables for syllable to spelling conversion, and 1,595 unique syllables and 1,239 unique phonemes for spelling to phoneme conversion. From these textbooks, we chose short sentences. This is because we intended to focus on students at Primary level and non-educational platform, language learners and unusual impaired persons with spelling style of Burmese primer. We prepared data to map syllable to spelling and spelling to phoneme with Perl scripts, but phonemes that have the same syllable and different pronunciations depending on their surrounding words are modified by manual checking. As an example, the syllable ဆည်း has two different phonemes - hse: for ဆည်းလည်း ("small bell" in English) and hsi: for ဆည်းပူး ("learn" in English). For syllable to spelling conversion, words are broken into syllables by *sylbreak* Perl script [12] and these syllables are mapped into spelling (example, က to ကကြီး-က). For spelling to phoneme conversion, spelled words are also broken into syllables and these syllables are mapped to their phonemes (example, က ကြီး က to ka. gyi: ka.). We currently referred the G2P mapping proposed by [13] why we need to study and understand the usage of IPA symbols. We randomized the data, then evaluated with 6-fold cross validation with about 1,160 sentences for open data set and one-tenth sentences from the training data for the closed data set.

B. Software

The following open source G2P converters, software frameworks and systems were applied for our G2P experiments:

1) Averaged Perceptron Tagger: It is implemented about 200 lines of python for tagging process. In this algorithm, features are sensitive to case (all words are lower case) and punctuation. The Viterbi algorithm is applied to find out the best pair of grapheme of input sentence and phoneme of output sentence for each input using current weight coefficients. Data preparation is similar to CRFsuite.
2) RDRPOSTagger: is a robust and easy-to-use toolkit for POS and morphological tagging. It employs an error-driven methodology to automatically construct tagging rules in the form of a binary tree [23]. It is a fast training and speed. It constructs a single classification ripple down rules tree to transform rules for POS tagging task. The RDRPOSTagger achieves very competitive accuracy measure in comparison of the state-of-the-art measure. Initial tagging of word/tag pair separated by white space character between each pair. The training part has been implemented by Python and the tagging process is both Python and Java.
3) CRFsuite: enhance the CRF models as fast as possible and an arbitrary number of features for each attribute which is not possible in CRF++ can be used. Its speed is faster than other CRF toolkit.

C. Evaluation

We used two evaluation criteria for G2P conversion output with 6-fold cross validation (i.e. in total 7,000 sentences corpus, about 5,800 for training and about 1,160 sentences for test data). One is BLEU (Bilingual Evaluation Understudy) and the other is $chrF^{++}$ (character n-gram F-score). In this study, two variants of $chrF^{++}$ score are reported – overall document level (F2), and macro averaged document level F-score (avgF2), which is the arithmetic average of sentence level scores [24]. BLEU measures the correspondence between output from the trained model (hypothesized text) and the reference text. BLEU is n-gram matches which is basically averaged.

TABLE II: BLEU and $chrF^{++}$ scores of Syllable to Spelling Conversion

Closed Data-set	BLEU(%)	$chrF^{++}$(%)	
		F2	avgF2
Averaged Perceptron	**100**	**100**	99.96
RDR	**100**	**100**	**99.98**
CRF	99.96	99.97	99.82

TABLE III: BLEU and $chrF^{++}$ scores of Syllable to Spelling Conversion

Open data-set	BLEU(%)	$chrF^{++}$(%)	
		F2	avgF2
Averaged Perceptron	97.52	98.15	97.66
RDR	**98.98**	**98.87**	**98.70**
CRF	96.65	96.84	96.00

TABLE IV: BLEU and $chrF^{++}$ scores of Spelling to Phoneme Conversion

Closed Data-set	BLEU(%)	$chrF^{++}$(%)	
		F2	avgF2
Averaged Perceptron	99.87	99.92	99.91
RDR	99.99	99.98	99.99
CRF	**100**	**100**	**100**

TABLE V: BLEU and $chrF^{++}$ scores of Spelling to Phoneme Conversion

Open data-set	BLEU(%)	$chrF^{++}$(%)	
		F2	avgF2
Averaged Perceptron	99.63	99.75	99.72
RDR	**99.91**	**99.91**	**99.90**
CRF	99.77	99.81	99.79

In other words, for each i-gram where i= 1,2,...N, the percentage of the i-gram tuples in the hypothesized text that also occur in the referenced text is computed. $chrF^{++}$ is an automatic evaluation tool for the output based on character n-gram precision and recall enhanced with word n-grams.

The BLEU and $chrF^{++}$ score results for G2P conversions experiments are shown in Table II (syllable to spelling on closed data-set), Table III (syllable to spelling on open data-set), Table IV (spelling to phoneme on closed data-set) and Table V (spelling to phoneme on open data-set). The numbers in bold indicate the highest score among the three conversion methods. Averaged Perceptron and RDR achieved same BLEU and $chrF^{++}$ scores for syllable to spelling conversion in Table II. Then RDR achieved the highest scores for all evaluation metrics as shown in Table III. For spelling to phoneme conversion at Table IV, CRF gained 100% accuracy on both BLEU and $chrF^{++}$ scores. In Table V, RDR achieved highest scores for both BLEU and $chrF^{++}$ scores. Moreover, the other two methods also gained the comparable results with the RDR.

VI. Error Analysis

We evaluated the errors occurred on the two G2P conversion results using PER measure with SCLITE (score speech recognition system output) program from the NIST scoring toolkit. SCLITE scoring and evaluating for the output of this experiment by comparing the reference text (human written out) to the hypothesized text (output from the trained model). Distance in words between the hypothesized text and the reference text is calculated by the Minimum Edit Distance (Levenshtein distance function) which estimates the cost of correct words. The formula of PER is as shown in Equation (3):

$$PER = (I + D + S) * 100/N \quad (3)$$

where I, S and D mean insertion, substitution and deletion of syllable and N means the total number of syllables. In our experiment, G2P models are trained with syllable segmented sentences and thus alignment was done on syllable units and the PER was derived from the Levenshtein distance at the phoneme level rather than the word level [2]. In our problem, only substitution case was occurred and any syllable were not deleted and inserted on all three methods. Phoneme level of syllable alignment in RDR is as follows:
syllable to spelling (bold words are substitutions by the RDR):
Scores: (#C #S #D #I) 2 1 0 0
REF: ဗထက်ခြိုက်-လုံးကြီးတင်-တစ်ချောင်းငင်-ကသတ်-ဗိုက် **သဝေထိုး-အ-ရေးချ-ငသတ်-အောက်ကမြစ်-အောင့်** တဝမ်းပူ-ယပက်လက်သတ်-တယ် ("stomach-ache" in English)
HYP: ဗထက်ခြိုက်-လုံးကြီးတင်-တစ်ချောင်းငင်-ကသတ်-ဗိုက် **ထဆင်ထူး-ရေးချ-ဝစ်စနှစ်လုံးပေါက်-ထား** တဝမ်းပူ-ယပက်လက်သတ်-တယ်

spelling to phoneme (phonemes substituted by RDR are capitalized):
For Burmese word (အင်္ဂါ ဗုဒ္ဓဟူး ("Tuesday and Wednesday" in English)) for spelling to phoneme conversion
Scores: (#C #S #D #I) 34 2 0 0
REF: a. jei a. ga. nge kin: zi: jei: cha. ga IN ga ba. de' chai' ta- chaun: ngin da. dwei: tha' bou' da. au' chai' da. ha. nha- chaun: ngin wi' sa. nha- loun: pau' HU:
HYP: a. jei a. ga. nge kin: zi: jei: cha. ga THEIN ga ba. de' chai' ta- chaun: ngin da. dwei: tha' bou' da. au' chai' da. ha. nha- chaun: ngin wi' sa. nha- loun: pau' LA:

Table VI and VII show that RDR is the lowest error rate relative to other two methods.

TABLE VI: PER of Syllable to Spelling Conversion

Avg Perceptron	RDR	CRF
2.4(%)	**1.0(%)**	2.7(%)

TABLE VII: PER of Spelling to Phoneme Conversion

Avg Perceptron	RDR	CRF
0.2(%)	**0.1(%)**	0.2(%)

VII. Result

Table II to Table V present the accuracy result of our grapheme to phoneme conversion. It is clear that the RDR is the best evaluation method for syllable to spelling conversion on both closed and open data-set, and spelling to phoneme conversion on open data-set. CRF completely achieved on spelling to phoneme conversion on closed data-set. In our experiment, The 6 fold incremental training/testing/evaluation was done for Averaged Perceptron, RDR and CRF approaches There is little significant on closed test. Therefore, the results for open data-set are shown in Figure 3 and Figure 4.

In these two figures, "ap" means Averaged Perceptron. In Figure 3, Averaged Perceptron and CRF dramatically increase BLEU scores rate from 1K to 3K training sentences. Then the BLEU scores of Averaged Perceptron and CRF reached 97.52 and 96.65 respectively. The best BLEU score was achieved by RDR. In Figure 4, RDR gradually increased the BLEU score from 99.35 to 99.91. CRF and Averaged Perceptron also gained the comparable results with the RDR. In these two figures, RDR is the best conversion method for both conversions on open data-set. It is obviously that the BLEU scores of Averraged Perceptron and CRF drop sharply when the training data are very low. And as we increase data size,

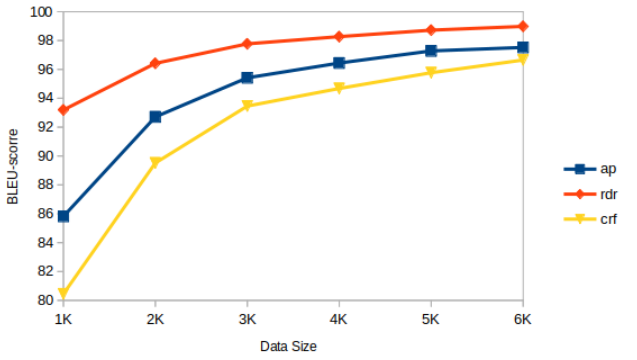

Fig. 3: BLEU Scores with Averaged Perceptron, RDR and CRF on varying data set sizes Syllable to Spelling Conversion

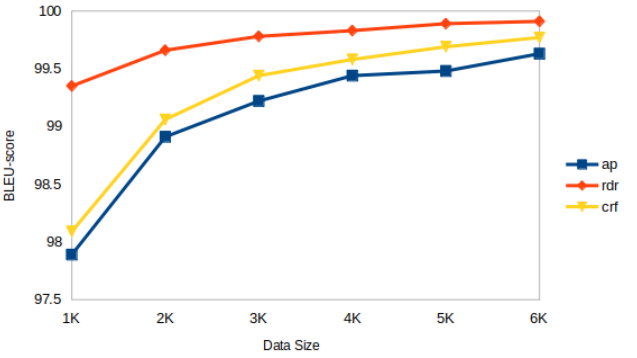

Fig. 4: BLEU Scores with Averaged Perceptron, RDR and CRF on varying data set sizes Spelling to Phoneme Conversion

TABLE VIII: Prediction Errors of Syllable to Spelling Conversion

Method	Ref.	Incorrect result (hypothesis)	Description
Avg Perceptron	မ-ရပင့်-ရေးချ-မျာ	မ-ရေးချ-ဝစ်စနှစ်လုံးပေါက်-မား	OOV error
RDR	မ-ရပင့်-ရေးချ-မျာ	ထဆင်ထူး-ရေးချ-ဝစ်စနှစ်လုံးပေါက်-ထား	RDR replaces all OOV words with ထဆင်ထူး-ရေးချ-ဝစ်စနှစ်လုံးပေါက်-ထား
CRF	မ-ရပင့်-ရေးချ-မျာ	ဗထက်ခြိုက်-ရပင့်-ရေးချာ-ဗျာ	OOV
Avg Perceptron	ကကြီး-ရရစ်-ယပက်လက်သတ်-ကြယ်	ခခွေး-ရရစ်-ယပက်လက်သတ်-ခြယ်	error in word ကကြီး
Avg Perceptron	ဃကြီး-ရေးချ-ဃာ	ဇကွဲ-ယပက်လက်သတ်-ဇယ်	hypothesis is far from the correct spelling
CRF	ဂငယ်-နှစ်ချောင်းငင်-ဂူ	ခခွေး-ရရစ်-သေးသေးတင်-ခြံ	hypothesis is far from the correct spelling

TABLE IX: Prediction Errors of Spelling to Phoneme Conversion

Method	Ref.	Incorrect result (hypothesis)	Description
Avg Perceptron	nja. nga. tha' njin **njin**	nja. nga. tha' njoun **njoun**	consonant error in njin
RDR	nja. nga. tha' **njin**	nja. nga. tha' **jin**	consonant error in njin
CRF	nja. nga. tha' **njin**	nja. nga. tha' **mjin**	consonant error in njin
Avg Perceptron	na. nge ha. htou: sa. tha' **nhi'**	na. nge ha. htou: sa. tha' **nha-**	tone error in nhi'
RDR	na. nge ha. htou: sa. tha' **nhi'**	na. nge ha. htou: sa. tha' **nha-**	tone error in nhi'
CRF	ga. nge nha-chaun: ngin **gu**	ga. nge nha-chaun: ngin **htu**	consonant error in gu

accuracy increased and error rate decreased.

When we also investigated errors for both conversions, we found that CRF has the two phoneme errors occurred in the Pali-loan word, and Averaged Perceptron and RDR are error free on the closed data-set for syllable to spelling conversion. These two errors are:

Ref: သ-တသတ်-သတ်-တဝမ်းပူ-တစ်ချောင်းငင်-တု
CRF: အက်ခရာ-၏ and
Ref: ဗထက်ခြိုက်-တစ်ချောင်းငင်-ဒဋ္ဌေးသတ်-ဗုဒ်-ဓအောက်ခြိုက်-သေးသေးတင်-ခံ
CRF: သ-ညသတ်-သည်

Then Averaged Perceptron got 0.1% of PER and RDR had 0.5% of sentences with errors on the closed test of spelling to phoneme conversion. On the open data-set of both conversions, most of the common errors are OOV and others are errors that produce the incorrect hypothesis from the reference. Except for OOV case, the RDR approach predict completely for syllable to spelling conversion. Example mostly errors occurred at syllable to spelling conversion is in Table VIII. For spelling to phoneme conversion, there are errors that produce the incorrect hypothesis from the reference when one reference grapheme has different hypothesized pronunciations on all three methods and most OOV errors. Example errors of spelling to phoneme conversion are shown in Table IX.

VIII. Conclusion

This experiment aims to measure the performance of spelling Burmese G2P conversion by comparing different machine learning methods. The automatic evaluation (PER) showed that RDR hit the lowest error rate on both conversions and CRF completely achieved with 100% accuracy of closed test on spelling to phoneme conversion when applied to spelling Burmese primer prediction. RDR is the best method giving the good prediction result of the syllable to spelling conversion except for OOV and also obtained the highest accuracy relative to the two other methods on spelling to phoneme conversion. Although the

Averaged Perceptron and CRF are not the best methods in our experiment, their open test evaluation scores are over 96% on syllable to spelling conversion and 99% spelling to phoneme conversion. Our experiment focuses on the spelled out of syllable without considering reading backward words. For example, အစ်ကိုကြီး (elder brother in English) is continuously spelled out like အ စသတ် အစ် ကကြီး လုံးကြီးတင် တစ်ချောင်းငင် ကို ကကြီး ရရစ် လုံးကြီးတင် ဆန်ခတ် ဝစ်စနှစ်လုံးပေါက် ကြီး. In future, we plan to test our G2P conversion by reading backward words (at a phrase-level) that have been read. For example, အစ်ကိုကြီး is spelled out like အ စသတ် အစ် ကကြီး လုံးကြီးတင် တစ်ချောင်းငင် ကို "အစ်ကို" ကကြီး ရရစ် လုံးကြီးတင် ဆန်ခတ် ဝစ်စနှစ်လုံးပေါက် ကြီး "အစ်ကိုကြီး".

References

[1] Chaw Su Hlaing and Aye Thida, "Phoneme based Myanmar text to speech system," International Journal of Advanced Computer Research - Vol 8(34), 2018, pp. 47–58.

[2] 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, 2016, pp. 11–22.

[3] Ye Kyaw Thu, Win Pa Pa, Andrew Finch, Aye Mya Hlaing, Hay Mar Soe Naing, Eiichiro Sumita and Chiori Hori, "Syllable Pronunciation Features for Myanmar Grapheme to Phoneme Conversion," In Proceedings of the 13th International Conference on Computer Applications (ICCA 2015), February 5 6, 2015, Yangon, Myanmar, pp. 161-167.

[4] Ritu Banga and Pulkit Mehndiratta, "Tagging Efficiency Analysis on Part of Speech Taggers," 2017 International Conference on Information Technology, 2017, pp. 264–267.

[5] Wint Theingi Zaw, Shwe Sin Moe, Ye Kyaw Thu, Nyein Nyein Oo, "Applying Weighted Finite State Transducers and Ripple Down Rules for Myanmar Name Romanization," In Proceedings of the 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON 2020), June 24-27, 2020, Virtual Conference Hosted by College of Computing, Prince of Songkla University, Thailand, pp. 143-148.

[6] Khaing Hsu Wai, Ye Kyaw Thu, Swe Zin Moe, Hnin Aye Thant, Thepchai Supnithi, "Myanmar (Burmese) String Similarity Measures based on Phoneme Similarity," Journal of Intelligent Informatics and Smart Technology, April 1st Issue, 2020, pp. 27-34.

[7] Denis Jouvet, Dominique Fohr and Irina Illina, "Evaluating grapheme-to-phoneme converters in automatic speech recognition context," ICASSP - 2012 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Mar 2012, Kyoto, Japan. pp.4821- 4824.

[8] Josef R. Novak, Nobuaki Minematsu and Keikichi Hirose, "WFST-based Grapheme-to-Phoneme Conversion: Open Source Tools for Alignment, Model-Building and Decoding," Proceedings of the 10th International Workshop on Finite State Methods and Natural Language Processing, pp 45–49.

[9] Shammur Absar Chowdhury, Firoj Alam, Naira Khan and Sheak R. H. Noori, "Bangla Grapheme to Phoneme Conversion Using Conditional Random Fields," 2017 20th International Conference of Computer and Information Technology (ICCIT), 22-24 December, 2017.

[10] Sittipong Saychum, Sarawoot Kongyoung, Anocha Rugchatjaroen, Patcharika Chootrakoo, Sawit Kasuriya and Chai Wutiwiwatchai, "Efficient Thai Grapheme-to-Phoneme Conversion Using CRF-Based Joint Sequence Modeling," INTERSPEECH 2016, September 8–12, 2016, San Francisco, USA. pp 1462-1466.

[11] Stanley F. Chen, "Conditional and joint models for grapheme-to-phoneme conversion," Eurospeech 2003, pp. 2033-2036.

[12] https://github.com/ye-kyaw-thu/sylbreak.

[13] Ye Kyaw Thu, Win Pa Pa, Andrew Finch, Jinfu Ni, Eiichiro Sumita and Chiori Hori, "The Application of Phrase Based Statistical Machine Translation Techniques to Myanmar Grapheme to Phoneme Conversion," In Proceedings of the Pacific Association for Computational Linguistics Conference (PACLING 2015), May 19 21, 2015, Legian, Bali, Indonesia, pp. 170-176.

[14] Ziga Golob, Jerneja Aganec Gros, Mario Zganec, Bostjan Vesnicer and Simon Dobrisek, "FST-Based Pronunciation Lexicon Compression for Speech Engines," Internationnal Journal of Advancced Robotic Systems, Aug 2012, pp. 1-8.

[15] Ye Kyaw Thu, Win Pa Pa, Jinfu Ni, Yoshinori Shiga, Andrew Finch, Chiori Hori, Hisashi Kawai, Eiichiro Sumita, "HMM Based Myanmar Text to Speech System," In Proceedings of the 16th Annual Conference of the International Speech Communication Association (INTERSPEECH 2015), September 6-10, 2015, Dresden, Germany, pp. 2237-2241.

[16] Vichet Chea, Ye Kyaw Thu, Chenchen Ding, Masao Utiyama, Andrew Finch and Eiichiro Sumita, "Khmer Word Segmentation Using Conditional Random Fields," In Proceedings of the 16th Annual Conference of the International Speech Communication Association (INTERSPEECH 2015), In Khmer Natural Language Processing 2015 (KNLP2015), December 4, 2015, Phnom Penh, Cambodia.

[17] Aye Mya Hlaing, Win Pa Pa and Ye Kyaw Thu, "Myanmar Number Normalization for Text-to-Speech," Proceedings of PACLING 2017, August 16-18, 2017, Yangon, Myanmar, pp. 346-356.

[18] Vathnak Sar and Tien-Ping Tan, "Applying Linguistic G2P Knowledge on a Statistical Grapheme-to-Phoneme Conversion in Khmer," The Fifth Information Systems International Conference, 2019, pp. 415–423.

[19] Hrafn Loftsson and Robert Ostling, "Tagging a Morphologically Complex Language Using an Averaged Perceptron Tagger: The Case of Icelandic," Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013); Linkoping Electronic Conference Proceedings, pp 105-119. 105-119. 105-119.

[20] J. Votrubec, "Morphological Tagging Based on Averaged Perceptron," WDS'06 Proceedings of Contributed Papers, 2006,pp. 191-195.

[21] https://explosion.ai/blog/part-of-speech-pos-tagger-in-python.

[22] Drahomira Johanka Spoustova, Jan Hajic, Jan Raab and Miroslav Spousta, "Semi-supervised Training for the Averaged Perceptron POS Tagger," Proceedings of the 12th Conference of the European Chapter of the ACL, pp. 763–771.

[23] http://rdrpostagger.sourceforge.net.

[24] Aye Mya Hlaing, Win Pa Pa and Ye Kyaw Thu, "Myanmar Number Normalization for Text-to-Speech," Proceedings of PACLING 2017, August 16-18, 2017, Yangon, Myanmar, pp. 346-356.

[25] Nguyen, Dat Quoc, Nguyen, Dai Quoc, Pham, Dang Duc, and Pham, Son Bao, "RDRPOSTagger: A Ripple Down Rules-based Part-Of-Speech Tagger", In Proceedings of the Demonstrations at the 14 th Conference of the European Chapter of the Association for Computational Linguistics, 2014, pp. 17–20.

[26] Nguyen, Dat Quoc, Nguyen, Dai Quoc, Pham, Dang Duc, and Pham, Son Bao. "A robust transformation-based learning approach using ripple down rules for part-of-speech tagging", AI communications 29.3 (2016), pp. 409-422.

[27] Rosenblatt, Frank, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain", Psychological Review 65 (6), 1958, pp. 386-408.

[28] Lafferty, J., McCallum, A., Pereira, F., "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", Proc. 18th International Conf. on Machine Learning. Morgan Kaufmann, 2001, pp. 282–289.

Hnin Yu Hlaing is a Tutor of Faculty of Information Science (FIS), University of Computer Studies, Meiktila and also a Ph.D candidate at University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myanmar. She got the master degree of Computer Science from University of Computer Studies, Mandalay, Myanmar. She is currently pursuing her Ph.D. studies in Grapheme to Syllable Sequence Phoneme conversion of spelling Myanmar language primer.

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.

Hlaing Myat Nwe is a PhD candidate of University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myanmar. A native of Myanmar, she holds a master degree of Information Science and Technology, and a bachelor degree of Information Science and Technology from University of Technology (Yatanarpon Cyber City), Myanmar. Her research interests include human-computer interaction, natural language processing and audio signal processing. She has been working to find efficient and user-friendly text input interfaces for Myanmar Sign Language. She is also a supervising lab members from NLP-Lab, UTYCC.

Hnin Aye Thant She is currently working as a Professor and Head of Department of Information Science at the University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin Township, Mandalay Division, Myanmar. She got Ph.D (IT) Degree from University of Computer Studies, Yangon, Myanmar in 2005. The current responsibilities are managing professional teachers, doing instructional designer of e-learning content development and teaching. She has 14 years teaching experiences in Information Technology specialized in Programming Languages (C,C++, Java and Assembly), Data Structure, Design and Analysis of Algorithms/Parallel Algorithms, Database Management System, Web Application Development, Operating System, Data Mining and Natural Language Processing. She is a member of research group in "Neural Network Machine Translation between Myanmar Sign Language to Myanmar Written Text" and Myanmar NLP Lab in UTYCC. She is also a Master Instructor and Coaching Expert of USAID COMET Mekong Learning Center. So, she has trained 190 Instructors from ten Technological Universities, twelve Computer Universities and UTYCC for Professional Development course to transform teacher-centered approach to learner-centered approach. This model is to reduce the skills gap between Universities and Industries and to fulfill the students' work-readiness skills.

Htet Ne Oo is currently working as an Associate Professor at Department of Information Science and Technology, University of Technology (Yatanarpon Cyber City), Myanmar. She had completed B.E (Information Technology) in 2007, M.E (Information Technology) in 2009. She also completed her Ph.D. in 2014 in Information Technology with specialization of Network security. She was a Research Fellow under Research Training Fellowship for Developing Country Scientists (RTFDCS) funded by DST, India in 2013. She also did online courses in ASEAN Cyber University Project in corporation with Busan Digital University. She has more than 11 years of teaching experience and 8 years of supervising bachelor, master and Ph.D. students. Her research fields include Geographic Information System, Data Mining, Machine Learning, Network Security, and Information Management.

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. Since2001, he has been with the Human Language Technology Laboratory, NECTEC, Thailand.

Aung Win graduated B.E from Yangon Technological University in 2002 and graduated M.E and Ph.D in 2005 and 2009 from Moscow Institute of Electronic Technology (MIET), Russia. All degree specialized in ICT and research conducted in Analog-Digital Computing System. He served in Ministry of Science and Technology and Ministry of Education as a Head of Academic Department, Head of Research Department, Principal of Technological University and currently serving as a Rector of University of Technology (Yatanarpon Cyber City) - UTYCC and can also call a founder of UTYCC. He also served as an Academic Leader for ICT area for developing, reviewing curriculum and syllabus for 33 Technological Universities from 2010 to 2014.