The State-of-the-Art Phonemic and Syllable Unit Transcriptions for Myanmar Names

1st Shwe Sin Moe
Department of Computer Engineering and Information Technology,
Yangon Technological University,
Myanmar,
shwesinmoe.ytu25@gmail.com

2nd Ye Kyaw Thu
National Electronics and Computer Technology Center (NECTEC),
Pathumthani, Thailand,
Language Understanding Lab.,
Myanmar,
yktnlp@gmail.com

3rd Wint Theingi Zaw
Department of Computer Engineering and Information Technology,
Yangon Technological University,
Myanmar,
wint.wtgz@gmail.com

4th Ni Htwe Aung
Department of Computer Engineering and Information Technology,
Yangon Technological University,
Myanmar,
nhadec@gmail.com

5th Nyein Nyein Oo
Department of Computer Engineering and Information Technology,
Yangon Technological University,
Myanmar,
nno2005@gmail.com

6th Thepchai Supnithi
National Electronics and Computer Technology Center (NECTEC),
Pathumthani, Thailand,
thepchai.supnithi@nectec.or.th

***Abstract*—Transcription is referred to the process of the systematic representation of a language in written form. The sources which can either be utterances (speech or sign language) or pre-existing text in another writing system are also included as a part of transcription. The process which represents speech by using a unique symbol for each phoneme of the language is called phonemic transcription. In this paper, we explore the proposed Long Short Term Memory (LSTM) based on Recurrent Neural Networks (RNNs) and the connectionist temporal classification (CTC) loss function to achieve the phonemic and syllable unit transcriptions of Myanmar names. We perform incremental experiments on total 7,000 speech and two types of label files (IPA phonetic transcription and syllable breaked Myanmar text) of Myanmar personal names. The performance of the transcription is evaluated based on the label error rate (LER) and word error rate (WER). The results show that the model can label more accurately on the phoneme transcription than the syllable unit transcription for Myanmar names.**

***Index Terms*—Phonemic transcription, Syllable unit labeling, Long Short Term Memory (LSTM) based on Recurrent Neural Networks (RNNs), Connectionist Temporal Classification (CTC), Myanmar personal names**

I. Introduction

Phonemic transcription, also known as phonetic script or notation, represents the speech which uses just a unique symbol for each phoneme of the language. Phonetic alphabets, such as the International Phonetic Alphabet (IPA) [9] is used for the most common type of phonemic transcription. Standard written form for some languages, such as English and Tibetan (including Myanmar language) is often irregular and difficult to predict pronunciation from spelling. The pronunciation of a one-to-one relationship between sound and symbol is made to be shown by phonemic transcription in most of the languages. Phonemic transcription identifies changes in pronunciation. The example for phonemic and syllable unit transcriptions of Myanmar names is shown in Figure 1. In Figure 1, for Myanmar name "ခင်ဇော်ခြည်နှိုင်" and its corresponding phoneme "kʰɪɴ za t͡ɕʰi naɪɴ", phonemic and syllable unit transcriptions are made according to their pronunciation. This paper

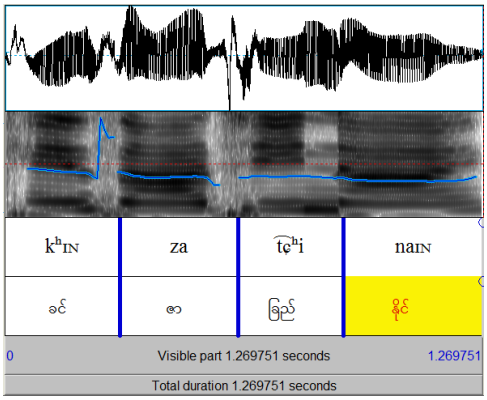

Fig. 1. A name from the Myanmar name Romanization corpus. From top to bottom: waveform of speech file; spectrogram with F0; phonemic transcription; Myanmar name

978-1-7281-9896-5/20/$31.00 ©2020 IEEE

investigates the phonemic transcription and syllable unit labeling of Myanmar names from speech based on connectionist temporal classification (CTC) and Long Short Term Memory (LSTM) based Recurrent Neural Networks (RNNs). One more contribution is we developed a 7K speech corpus for Myanmar personal names.

The structure of this paper can be described as follows: a brief review of phonemic transcription using Persephone tools [1], which is Long Short Term Memory (LSTM) based Recurrent Neural Networks (RNNs) and connectionist temporal classification (CTC) is presented in section II. In section III, we explain the nature of the Myanmar Language. Data Preparation for our experiments is described in section IV. Section V introduces the methods used to build the models briefly. In section VI, we present the experimental setup of this paper. Results and discussions of our experiments are described in section VII. The error analysis based on the experimental results is discussed in section VIII.

II. Related Work

Oliver Adams et al. [1] proposed an automatic phonemic transcription tool, named Persephone, which was based on LSTM based RNN and CTC. In that paper, the use of neural network architecture with the connectionist temporal classification loss function was presented to predict the direct labelling of phonemes and tones given an acoustic model in a language documentation setting. Experiments were made on two low-resource tonal languages (Yongning Na and Eastern Chatino) for this approach and the results were reported as the applicability and transcription was very encouraging and had been incorporated into the workflow for transcribing higher to untranscribed speech and reviewing existing transcriptions.

Alexis Michaud et al. [2] presented the automatic phonemic transcription method used for speech recognition. The paper illustrated qualitative error analysis on two tests (Yongning Na and Tsuut'ina). Error analysis allowed a renewed exploration of phonetic details such as examining the output of phonemic transcription compared with spectrographic and aural evidence. The method supported the automatic phonemic transcription systems such as Persephone by applying the LSTM based RNN and CTC. There was no conspicuous difference between function words and lexical words in terms of error rates in phonemic recognition for Na. The findings suggested that the gains in accuracy could be obtained by corporating word boundary information in the training set.

Guillaume Wisniewski et al. [3] described some possibilities for automating the process of data preprocessing steps in Persephone. The authors reported the tests aimed at investigating possibilities for automating the process of data preprocessing and used the same dataset from the Yongning Na language as was used in a previous study [1]. The results stated that Persephone achieved better results for predicting phonological transcriptions with phonemes than for predicting orthographic transcriptions with phonemes by using the same data.

David R. Mortensen et al. [4] proposed Epitran, a system for G2P (grapheme-to-phoneme) mapping which supported 61 languages. It mapped the orthographic data into the phonetic space and produced phonemic representation in either International Phonetic Alphabet (IPA) [9] or X-SAMPA as outputs. It became a useful resource for researchers who were working with speech, signal processing, and Natural Language Processing as the case with high-quality modes were added for new languages.

Ye Kyaw Thu et al. [15] presented the first mapping of Grapheme to Phoneme conversion for Myanmar language. The authors proposed four simple Myanmar syllable pronunciation patterns as features that can be used to augment the models in a Conditional Random Field (CRF) approach to G2P conversion. The Myanmar Language Commission (MLC) Pronunciation Dictionary was used as a basis for pronunciation mapping. The results of G2P experiments based on the Myanmar Language Commission (MLC) Pronunciation Dictionary by using CRF for Myanmar Language were also described in this paper.The results showed that the new features can substantially improve the accuracy of grapheme to phoneme conversion.

Ye Kyaw Thu et al. [16] introduced the comparison of grapheme–to–phoneme conversion methods on a Myanmar Pronunciation Dictionary. The authors examined seven G2P conversion approaches: Adaptive Regularization of Weight Vectors (AROW) based structured learning (S-AROW), Conditional Random Field (CRF),Joint-sequence models (JSM), phrase-based statistical machine translation (PBSMT),Recurrent Neural Network (RNN), Support Vector Machine (SVM) based point-wiseclassification, Weighted Finite-state Transducers (WFST) on a manually tagged Myanmar phoneme dictionary for for incremental training with a small Myanmar language G2P lexicon. The G2P bootstrapping experimental results were measured with both automatic phoneme error rate (PER) calculation and manual checking in terms of voiced/unvoiced, tones, consonant and vowel errors. The results showed that the CRF, Phonetisaurus, and SMT approaches gave rise to the the lowest error rates on the most important features of Myanmar G2P conversion: voiced/unvoiced, vowel patterns and tone.

Based on the experiments made in the previous works, the experiments were carried out to study the performance of phonemic and syllable unit transcriptions for Myanmar names in this paper.

III. Nature of Myanmar Language

Myanmar Language, formerly known as Burmese, which is included in Sino-Tibetan Language family, is a tonal, syllable-timed language and largely monosyllabic, and analytic language. It is an official language of Bamar people in Myanmar and is spoken by two-thirds of population in Myanmar. In Myanmar language, 12 vowels, 33

consonants, and 4 medials are used as basics alphabets. Myanmar syllables or words are basically constructed by using the combination of consonants and vowels. Four nominal tones such as low, high, creaky and checked can be found in Myanmar Language and the different tones of Myanmar words have different meanings.

IV. Data Preparation

This section explains how to prepare text and speech data for Myanmar Names. Currently, there is no freely available Myanmar name Romanization corpus. And thus we are developing a corpus for Myanmar names including city, food, organization names, etc [5]. Firstly, we collect the names of people from social media such as Facebook and student affairs from our university. We also collect the Myanmar names of city, food, and organization from Wikipedia. We consider various local Romanization of Myanmar people for real-word NLP applications. The speech data is collected by recording Myanmar names from the developing names Romanization corpus. However, in this paper, we used only personal names for the transcription experiments.

V. Method

A Long Short Term Memory (LSTM) based on Recurrent Neural Network (RNNs) in a bidirectional configuration is used as the underlying model in our system [1]. Long Short Term Memory (LSTM) is a special part of neural networks called Recurrent Neural Networks (RNNs) which are capable of processing sequential input of unbounded and arbitrary length. It is suitable for the labeling task. A hidden layer in LSTM consists of a set of recurrently connected blocks, also known as memory blocks. Each of the memory blocks contains one or more recurrently connected memory cells and three multiplicative units the input, output, and forget gates that provide to write, read, and reset operations for the cells. The input to the cells is multiplied by the activation of the input gate, the output to the net is multiplied by the output gate, and the previous cell values are multiplied with the forget gate. The net can only interact with the cells via the gates [8]. The LSTM in a bidirectional configuration (BLSTM) [11] has two separate recurrent hidden layers, and both of them are connected to the same input and output layers. The BLSTM recurrent neural network is trained with the CTC algorithm [12], like gradient descent, combined with backpropagation through time to compute the gradients needed during the optimization process, in order to change each weight of the LSTM network in proportion to the derivative of the error (at the output layer of the LSTM network) with respect to corresponding weight. [13]

Connectionist Temporal Classification (CTC) allows RNN to be trained on unsegmented sequence data. CTC refers to the outputs and scoring, and is independent of the underlying neural network structure. The input is a sequence of observations and the outputs are a sequence of labels, which can include blank outputs. [14] Interpreting the network outputs as a probability distribution over all possible label sequences, conditioned on a given input sequence is the basic idea of CTC. Given this probability distribution, an objective function can be derived from the probabilities of the correct labeling. The objective function is differentiable and thus the network can be trained with standard backpropagation through time [6]. Alignments between speech frames and labels in the transcription are made by training with CTC. The use of an underlying recurrent neural network allows the model to implicitly model context via the parameters of the BLSTM, despite the independent frame-wise label predictions of the CTC network. The network can then be used as a classifier by selecting the most probable labeling for a given input sequence [10].

A CTC network has a softmax output layer with one more unit than there are labels in L [10]. The activations of the first |L| units are interpreted as the probabilities of observing the corresponding labels at particular times. The probability of observing a 'blank' or no label is the activation of the extra unit. The total probability of anyone label sequence can then be found by summing the probabilities of its different alignments [10].

VI. Experimental Setup

A. Corpora Statistics

The Myanmar name speech corpus is built by recording the Myanmar names in the text corpus. The corpus is recorded with only one female speaker. "TASCAM DR-44WL" Recorder is used and the speech files are recorded in the setting of wave file (.wav) format and the sampling rate of 48 kHz with the mono channel. The duration of each speech file is between 0 sec and 4 sec.

We used 7K names including not only Bamar names but also ethnic names of our country such as Kachin, Kayah, Kayin, Shan, Chin, and Rakhine. The transcription experiments are run on the datasets ranging from 1K to 7K personal names in size and the evaluation is done against the training set size. The performance evaluation is mainly measured with label error rate (LER) and accuracy and we also did error analysis in terms of word error rate (WER).

B. Workflow

The syllable segmentation was done on manually prepared Myanmar personal names. "sylbreak" syllable segmentation tool based on RE (Regular Expression) was used (source code link: https://github.com/ye-kyaw-thu/sylbreak). The syllable breaked personal names were converted into their corresponding phonemes by using "Epitran" [4], a massively multilingual G2P system. These two types of labels (1) syllable segmented Myanmar names and (2) syllable segmented phonemes were used for building transcription models. The format of the two types of label files were prepared as the following:

(1) ကောင်း ကောင်း စံ (syllable segmented name)
(2) kaun kaun san (syllable segmented phoneme)

We did the incremental training (from 1K to 7K) with Persephone [1]. It can be used to help bring transcriptions closer to the audio.

Generally, the transcription tool can be run over the training dataset, finding instances where the probability for the transcription is low, and flagging these for manual verification. Alternatively, the user can measure the edit distance between the predicted transcript (hypothesis) and the manual transcript (reference).

As the first step, the log filterbank features for frames in the audio are computed if the filename of (.wav) extension files is found in the folder "wav". These features can be considered as a sort of spectrogram representing the energy at different frequencies throughout the recording. For each (.wav) file, there can be found corresponding transcription file with label type (with extension .phonemes) and text file (with extension .txt) in "label" folder. For evaluation of the model, if the data is found in the required format, wave normalization and speech feature extraction will be divided into training, validation, and test sets in the ratio of 95:5:5 [1]. These are typically and randomly selected, either at the name level or the utterance level.

The core of the system is LSTM based RNNs and CTC loss function that takes the filter bank features as input and produces a phonemic transcription of the entire utterance. We trained each configuration with a minimum of 30 epochs. Training will continue until the maximum LER for the validation set is met or another stopping condition occurs. Batch size must be between 4 and 64 utterances because it varies according to the training set size. Data are trained in the setting of 3 hidden layers with 250 hidden units. Next, the model learned to provide label predictions for each of the frames. These labels might be phonemes, tones, or orthographic characters but in our experiment, the labels might be syllable breaked Myanmar names and phonemes. After we conducted the experiments, we got the output labelled files which show how the models can make automatic labeling of speech with phonemes and syllable breaked Myanmar names correctly.

C. Tools used for Experiments

- Epitran: Epitran, version 1.3, (source code link: https://github.com/dmort27/epitran), is a python library and tool which is based on python for transliterating orthographic text IPA (International Phonetic Alphabet) [4].
- Persephone: Persephone, v0.4.2 (beta version), (source code link: https://github.com/persephone-tools/persephone), is an automatic phoneme transcription tool which is implemented is an Python/Tensorflow with extensibility, bidirectional Long Short Term Memory (BLSTMs) and the connectionist temporal classification (CTC) loss function. It is designed for situations where training data is limited, such as an hour of transcribed speech and it is possible to use small amounts of data to train a transcription model [1].

D. Evaluation

We evaluate the performance of our system by computing the average edit distance between the predicted (hypothesis) and manual (reference) transcripts or labels of the test set (i.e. the label error rate LER). The label error rate (LER) is calculated based on the number of insertions, substitutions, and deletions according to the following equation:

$$LER = \frac{(D - S - I)}{N} \tag{1}$$

where D is the total number of deletions in the hypothesis (hyp), S is the total number of substitutions, I is the total number of insertions, and N is the total number of labels in the references [6]. The lower the label error rate (LER), the better the model can label correctly speech with phonemes and syllable breaked Myanmar names.

E. Word Error Rate

We analyzed labelled outputs with word error rate (WER) by using the "SCTK" toolkit for making alignments between the manual transcript (reference) and the predicted transcript (hypothesis) [7]. The word error rate (WER) is computed by using the following equation:

$$WER = \frac{(I + D + S) \times 100}{N} \tag{2}$$

where I is the number of insertions, S is the number of substitutions, D is the number of deletions, N is the number of words in the reference and C is the number of correct words in the reference file ($N = S + D + C$). There is a point to note that the WER can be greater than 100% if the number of insertions is very high.

VII. Results and Discussions

The label error rate (LER) results for state-of-the-art phonemic and syllable unit transcriptions of Myanmar names are shown in Table I. Bold numbers in Table I indicate the best LER results of our experiments. The lower the label error rate (LER) and the , the better the performance of the system is. When the data gradually increase for the experiments, the LER results can be achieved better. The label error rates (LER) are decreased logarithmically according to the training dataset size. For our experiments, it can be seen clearly that the better LER results achieved by increasing data gradually in both of the phonemic and syllable unit transcriptions with Myanmar names. When we compare the LER results of these experiments, the results show that the model can label more accurately and correctly on phonemic

transcription. Therefore, we can assume that the state-of-the-art phonemic transcription has better performance (around LER of 0.08% and accuracy of 7.8%) than that of syllable unit transcription for Myanmar names. This indicate that phonemes transcription is easier than Myanmar syllable unit labeling for our model. We can increase the data size to get better performance in both conditions of transcription.

TABLE I
Label Error Rates (LER) for phonemes and syllable transcriptions of Myanmar names (lower is better)

Data	LER of phoneme transcription	LER of syllable transcription with Myanmar Names
1,000	0.686275	0.696078
2,000	0.359406	0.571287
3,000	0.350773	0.317108
4,000	0.317579	0.308624
5,000	0.289973	0.298845
6,000	0.221705	0.298616
7,000	**0.214198**	**0.298481**

VIII. Error Analysis

The word error rates (WER) for each experiment are calculated by using "SCLITE" program based on "SCTK" toolkit [7] and Equation (2). The results of WER for phonemic and the syllable unit transcriptions of Myanmar names are shown in Table II. The lowest word error rates (WER) are described with bold letters in Table II. From Table II, we found that the WER is slightly decreased when the data is increased gradually for both phonemic transcription and syllable unit transcription of Myanmar names with data 1K to 7K. The accuracies got by calculating the word error rates (WER) by using "SCTK" toolkit [7] are also shown in Table III. The best accuracies are described with bold numbers in Table III. The lower the WER result, the better the accuracy result is.

TABLE II
WER for phonemes and syllable transcriptions of Myanmar names (lower is better)

Data	WER of phoneme transcription	WER of syllable transcription with Myanmar Names
1000	70.7%	70.7%
2000	34.4%	56.6%
3000	35.4%	30.6%
4000	30.8%	29.7%
5000	28.8%	29.5%
6000	22.3%	29.3%
7000	**21.3%**	**29.1%**

The confusion matrix and word error rates (WER) are described in details by using "SCLITE" program. The following are some example calculations of WER for some of the names in experiments. For example,

TABLE III
Accuracy for phonemes and syllable transcriptions of Myanmar names (higher is better)

Data	Accuracy of phoneme transcription	Accuracy of syllable transcription with Myanmar Names
1000	29.3%	29.3%
2000	65.6%	43.4%
3000	64.6%	69.4%
4000	69.2%	70.3%
5000	71.2%	70.5%
6000	77.7%	70.7%
7000	**78.7%**	**70.9%**

phonemic transcription for the phoneme "kaʊn θAn jɪn" ("ကောင်းသန်းရင်" in Myanmar Name) is compared to the manual transcript (reference) and the output of "SCLITE" program shows as the following:

Scores: (#C #S #D #I) 3 1 0 0
REF: kaʊn θAn jɪn
HYP: kaʊn θəWɛ jɪn
Eval: S

In this output, one substitution "θAn = = > θəWɛ" is occurred and C is 2, D is 0, I is 0 and N is 3. WER for this phoneme is equal to 33%.

Scores: (#C #S #D #I) 2 1 0 0
REF: ကောင်း သန်း ရင်
HYP: ကောင်း သိန်း ရင်
Eval: S

In this case, one substitution "သန်း ==> သိန်း" is happened and C is 2, D is 0, I is 0 and N is 3 and thus WER is equal to 33% for this. And we find that phonemic and syllable unit transcriptions of Myanmar names have the same WER percentage for this case.

The next example is phonemic transcription of phonemes which is compared to the manual transcript (reference) and the output is shown in the following:

Scores: (#C #S #D #I) 3 0 1 0
REF: kʰɪn NI ni wɪn
HYP: kʰɪn ** ni wɪn
Eval: D

One deletion "NI = = > **" is found in this case and C3, I=0, S=0 and N is 4. And thus, WER for this word is 25%.

Scores: (#C #S #D #I) 2 1 1 0
REF: ခင် နီ နီ ဝင်း
HYP: ခင် ****** နှင်း ဝင်း
Eval: D S

In this case, there is one deletion "ဦ ==> ******" and one substitution "ဦ ==> ʒင်း". C is 2 and N is 4. So, WER is 50%.

Another example is the phonemic transcription of "t͡ɕaʊʔ lɪɴ tʰəkɛʔ" ("ကျော် လင်း ထက်" in Myanmar name) is compared to the manual transcript (reference), and the output shows:

Scores: (#C #S #D #I) 3 0 0 0
REF: t͡ɕaʊʔ lɪɴ tʰəkɛʔ
HYP: t͡ɕaʊʔ lɪɴ tʰəkɛʔ
Eval:

There is no substitution, deletion and insertion and thus WER is 0% in this case.

Scores: (#C #S #D #I) 3 0 0 0
REF: ကျော် လင်း ထက်
HYP: ကျော် လင်း ထက်
Eval:

In this case, substitution, insertion and deletion are not found and N is 3. WER is 0%. Therefore, we find that the model can label correctly with this name for both phonemic and syllable transcriptions of Myanmar names.

After we made error analysis on confusion pairs of each model in details, we found that some of the error are due to the system results that is not phonologically well formed and the mapping of sound with phoneme and syllable of Myanmar names. The top 30 confusion pairs for state-of-the-art phonemic transcription of Myanmar names is shown in Table IV.

Here, the top 30 confusion pairs occurs due to the wrong mapping between voice and phonemes of the name. This kind of confusion pairs can be reduced by recording the sound more clearly and make preprocessing carefully for speech.

Table V shows the top 30 confusion pairs for state-of-the-art syllable unit transcription of Myanmar Names. This show that most errors are the same as phoneme transcription and due to the wrong mappings between voice and syllable breaked Myanmar names.

TABLE IV
Confusion Pairs for phonemes transcription of Myanmar names

Freq	Confusion Pair (REF ===> HYP)
4	sʰu = = > su
2	bəja = = > ja
2	di = = > t͡ɕi
2	jəj = = > jəja
2	maʊʔ = = > maʊɴ
2	me = = > mɪɴ
2	mojɛʔ = = > mo
2	nəwe = = > me
2	saɴ = = > sʰaɴ
2	t͡ɕaʊʔ = = > t͡ɕɔ
2	t͡ɕɔ = = > t͡ɕaʊʔ
2	θaɴ = = > θəkɛʔ
1	bo = = > mo
1	boʊɴ = = > m̥oʊɴ
1	bu = = > p̥ʰu
1	daɪɴ = = > laɪɴ
1	di = = > pe
1	dɪɴ = = > kʰɪɴ
1	d͡ʑa = = > t͡ɕɔ
1	d͡ʑeɪɴ = = > ɲanɛʔ
1	d͡ʑu = = > tɕu
1	heɪɴ = = > he
1	həkɛʔkʰɛʔ = = > kʰəkɛʔ
1	ja = = > m̥u
1	ja = = > ta
1	jaɪɴ = = > ja
1	je = = > le
1	ji = = > ʔi
1	jo = = > ju
1	jəj = = > zɪɴ

IX. Conclusion

In this paper, we presented the state-of-the-art phonemic and syllable unit transcriptions for Myanmar Names by applying LSTM based RNNs and CTC approach. We used 7K of speech and two types of label files for Myanmar names to label directly speech with text. According to the experimental results, we found that syllable unit transcription for Myanmar names is more difficult than phonemic transcription for Myanmar names. This paper also presented the top 30 confusion pairs of phonemic and syllable unit transcriptions for Myanmar names. In the future, we plan to make transcription experiments with city, food and organization names from the developing names Romanization corpus.

References

[1] Oliver Adams, Trevor, Graham Neubig, Hilaria Cruz, Steven Bird, Alexis Michaud, "Evaluating Phonemic Transcription of low-Resource Tonal Languages for Language Documentation", Proceedings of LERC 2018: 11th edition of the Language Resources and Evaluation Conference, 7-12 May 2018, Miyazaki (Japan).

[2] Alexis Michaud, Oliver Adams, Christopher Cox, Severine Guillaume, "Phonetic lessons from automatic phonemic transcription: preliminary reflections on Na (Sino-Tibetan) and Tsuut'ina (Dene) data", ICPhS XIX (19th International Congress of Phonetic Sciences), August 2019, Melbourne, Australia.

[3] Guillaume Wisniewski, Alexis Michaud, Severine Guillaume,"Phonemic Transcription of Low-Resource Languages: To What Extent can Processing be Automated?", Proceedings of the 1st Joint SLTU and CCURL Workshop (SLTU-CCURL 2020), Pages 306-315, Language Resources and Evaluation Conference (LERC 2020), Marseille, 11-16 May 2020.

[4] David R. Mortensen, Siddharth Dalmia, Patrick Littell, "Epitran : Precision G2P for Many Languages",Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), May 2018, Page 2710-2714.

[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", ECTI Conference 2020, Thailand.

TABLE V
CONFUSION PAIRS FOR SYLLABLE UNIT TRANSCRIPTION OF MYANMAR NAMES

Freq	Confusion Pair (REF ===> HYP)
5	ကျော့ ==> ကျော်
4	ဆု ==> စု
2	က ==> ဟ
2	ကြီး ==> ကြည်
2	စန်း ==> ဆန်း
2	ဇ ==> စ
2	ညို ==> မျိုး
2	ထွန်း ==> ဖန်း
2	ဖန်း ==> ပြီး
2	ဖြိုး ==> ဖြူ
2	မိုရ် ==> မိုး
2	မေ ==> မင်း
2	ယျ ==> ယှာ
2	ရည် ==> ရီ
2	လှိုင် ==> နိုင်
2	ဝင့် ==> ဝင်း
2	သဇ္ဈာ ==> ဇာ
2	သန်း ==> သဲ
1	ကင် ==> ခင်
1	ကင်း ==> ကေ
1	ကို ==> စု
1	ကိုး ==> ကို
1	ကူ ==> သူ
1	ကော် ==> ကို
1	ကံ့ ==> ကံ
1	ကျင် ==> ကျိန်
1	ကျင် ==> ထင်
1	ကျင်း ==> ကြီး
1	ကျန် ==> ကြယ်
1	ကျိန် ==> တင်

[6] Florian Eyben, Martin Wollmer, Bjorn Schuller, Alex Graves, "From Speech to Letters Using a Novel Neural Network Architecture for Grapheme Based ASR", Pages 376-380, ASRU 2009.

[7] (NIST) The National Institute of Standards and Technology, Speech recognition scoring Toolkit (sctk), version 2.4.10, 2015.

[8] Alex Graves, Santiago Fernandez, and Jurgen Schmidhuber, "Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition", ICANN 2005, LNCS 3697, pp. 799–804, 2005.

[9] International Phonetic Alphabets, https://en.wikipedia.org/wiki/International_Phonetic_Alphabet.

[10] Alex Graves, Santiago Fernandez, Faustino Gomez, Jurgen Schmidhuber, "Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks", Proceedings of the International Conference on Machine Learning, ICML, Pittsburgh, USA, June 2006.

[11] Alex Graves, Santiago Fern´Andez, and J¨Urgen Schmidhuber, "Bidirectional LSTM networks for improved phoneme classification and recogni-tion", Proceedings of the 2005 International Conference on Artificial Neural Networks, Warsaw, Poland, 2005.

[12] Santiago Fernandez, Alex Graves, Jurgen Schmidhuber, "Phoneme recognition in TIMIT with BLSTM-CTC", April 22, 2008.

[13] https://en.wikipedia.org/wiki/Long short term memory

[14] https://en.wikipedia.org/wiki/Connectionist temporal classification

[15] 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.

[16] Ye Kyaw Thu, Win Pa Pa, Yoshinori Sagisaka, Naoto Iwahashi, "Comparison of Grapheme–to–Phoneme Conversion Methods on a Myanmar Pronunciation Dictionary", In Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), COLING 2016, December 11-17, 2016, Osaka, Japan, pp. 11–22.