English-Myanmar (Burmese) Phrase-Based SMT with One-to-One and One-to-Multiple Translations Corpora

Honey Htun
Department of Computer Engineering and Information Technology,
Yangon Technological University,
Yangon, Myanmar,
honeyhtun.85@gmail.com

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

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

Thepchai Supnithi
National Electronics and Computer Technology Center (NECTEC),
Pathum Thani,Thailand,
thepchai@gmail.com

Abstract*— This paper contributes the first investigation of machine translation (MT) performance differences between Myanmar and English languages with the use of several possible Myanmar translations for the specific primary educational domain. We also developed both one-to-one and many Myanmar translations corpora (over 8K and 46K sentences) based on old and new English textbooks (including Grade 1 to 3) which are published by the Ministry of Education. Our developing parallel corpora were used for phrase-based statistical machine translation (PBSMT) which is the de facto standard of statistical machine translation. We measured machine translation performance differences among one-to-many English to Myanmar translation corpora. The differences range between 19.68 and 52.38 BLEU scores from English to Myanmar and between 50.17 and 75.12 BLEU scores from Myanmar to English translation. We expect this study can be applied in Myanmar-to-English automatic speech recognition (ASR) development for primary English textbooks. The main purpose is to translate primary English textbooks data correctly even if the children use in several Myanmar conversation styles.

Keywords*— Phrase-based Statistical Machine Translation (PBSMT), One-to-Many Parallel Corpus, Myanmar-English Machine Translation, Primary English Textbooks of Myanmar, Word Error Rate (WER).

I. INTRODUCTION

Nowadays, our Myanmar children in rural and less developing areas are facing many difficulties in education. One of the challenges in their educational lives is that English is taught as a second language from kindergarten. The main point for continuous education is to catch up with the knowledge of English since childhood. However, there are no enough of teaching staffs and teaching aid devices. Consequently, most of the primary students in these rural and less developing areas are weak in learning English. To address these challenges, the machine translation technologies can be applied to be more interest in their lessons and to support as learning assistance tool. Although English to Myanmar as in bi-directional translation systems have been successfully used in other domains such as travel, to the best of our knowledge, there are no prior researches targeting the application of machine translation to the education sector. This paper contributes to the first studying of the machine translation performance between Myanmar primary educational textbooks sentences and translated Myanmar sentences by applying statistical machine translation (SMT) approaches (especially PBSMT) with one-to-one and one-to-many translated parallel data. One more contribution is that we are developing a parallel corpus of Myanmar-English on the primary educational domain. We also consider one-to-many possible manual translations for English to Myanmar translation direction.

The structure of the paper is as follows. In the next section, we present a brief review of machine translation systems for Myanmar-English as bi-directional systems. Section III presents the current state of English education at Primary schools in Myanmar and section IV describes the Myanmar-English parallel corpus building for machine translation experiments. In section V, we describe PBSMT as the experimental methodology used in machine translation experiments. Section VI presents the statistical information of the corpus and the experimental settings. Section VII reports the experimental results with discussions and section VIII presents the error analysis on translated outputs. Finally, section IX presents conclusion and future work.

II. RELATED WORK

This section reviews the previous works in statistical machine translation between Myanmar and English languages. To date, there have been some studies on the SMT of Myanmar language.

Thet Thet Zin et al. (2011) [1] described statistical Myanmar phrase translation system with morphological analysis. Here, the experiments were conducted based on total data size of 13,042: 12,827 parallel sentences of which were training set and the rest of 215 were test set. And Bayes' rule was used to reformulate the translation probability for translating Myanmar phrases into English phrases. The evaluation criteria of machine translation were precision, recall and the F-measure. There were problems with many out of vocabulary (OOV) words such as proper noun, noun and verb phrases in the first baseline system. As the second step, the morphological analysis is applied on pre-processing phrase of translation process to address the above OOV problem. According to the results, the morphological analysis method achieved the good comparison with the baseline. However, there were still most errors of post-positional markers that made ambiguous meaning. Therefore, one way of helping that problem, part-of-speech (POS) tagging technique, was applied. Adding morphology and POS of Myanmar language to baseline system gave the best results and reduced OOV rates. But, there were still 95 errors in 215 tested sentences such as unknown foreign words, translation failure, segmentation error, detecting verb phrases error, untranslatable phrases and missing English particles.

Ye Kyaw Thu et al. (2016) [2] presented the first large-scale study of the translation of the Myanmar language. There were a total of 40 language pairs in the study that included languages both similar and fundamentally different from Myanmar. In this experiment, 457,249 sentences were used for training, 5,000 sentences for development and 3,000 sentences for evaluation. The results showed that the hierarchical phrase-based SMT (HPBSMT) [3] approach gave the highest translation quality in terms of both the BLEU [4] and RIBES scores [5].

Win Pa Pa et al. (2016) [6] presented the first comparative study of five major machine translation approaches applied to low-resource languages. PBSMT, HPBSMT, tree-to-string (T2S), string-to-tree (S2T) and Operation Sequence Model (OSM) translation methods were applied to the translation of limited quantities of travel domain data between English and {Thai, Laos, Myanmar} in both directions. Here, 20,000 sentences were used for training, 500 sentences for development and 300 sentences for evaluation. The experimental results indicated that in terms of adequacy (as measured by BLEU score), the PBSMT approach produced the highest quality translations. Here, the annotated tree is used only for the English language for S2T and T2S experiments. This is because there is no publicly available tree parser for Lao, Myanmar and Thai languages. From their RIBES scores, we noticed that OSM approach achieved the best machine translation performance for Myanmar to English translation.

Rui et al. had developed Neural Machine Translation (NMT) and PBSMT systems with pre-ordering for English-Myanmar in both translation directions. All the provided parallel data for all the targeted translation directions, including the training corpus “ALT” and “UCSY” and the “ALT” dev/test data: 226,500 sentences for training, 1,000 sentences for testing and 900 sentences for evaluation were used. The source English part was pre-ordered before being input into NMT and SMT systems. The results also confirmed the slight positive impact of using pre-ordering in English-Myanmar PBSMT [7].

Based on the experimental results of the previous works, in this paper, the PBSMT experiments were carried out to study the performance variations using PBSMT, one-to-one and many translations corpora.

III. ENGLISH EDUCATION AT PRIMARY SCHOOLS IN MYANMAR

In this section, we would like to present about English education at primary schools in Myanmar. In Myanmar’s educational sector, English subject is taught as a second language since primary level. It is very important to catch up with the knowledge of the next higher levels because almost next higher level textbooks are published in English. Nowadays, there are two types of primary level English textbooks in Myanmar such as old primary curriculum and new ones. In old ones, the lessons are in general teaching style and the students are weak in interesting and understanding as a consequence.

Today, a drastic education reform has been implementing by the Ministry of Education in Myanmar. And primary education reform is one of the important topics. In 2014, CREATE Project (The Project for Curriculum Reform for Primary Level of Basic Education in Myanmar) was launched for emerging new primary education textbooks, Teacher’s Guide, updating assessment. And this project introduced new primary education to in-service and pre-service teachers. CREATE Project is jointly organized by the Ministry of Education in Myanmar and Japan International Cooperation Agency.

From June 2017 to current 2020, new primary Grade 1, 2 and 3 English textbooks were introduced nationwide. New primary education textbooks are richer and made up of attractive contents that promote students’ active learning, considering diversification like gender, ethnicities, disabilities, many pictures and photos that stimulate students’ interests of learning, colorful however applying universal color style to be friendly for color-blinded students. There are 36 weeks per year and one period is taken in 40 minutes for all Grades [8]. Grade 1 English textbooks cover for alphabetical letters, numbers, short nouns and Grade 2 English textbooks cover for adjective, verb, greeting sentence and short sentence forms. In Grade 3, long sentences, alternate practice sentences, and usage of question words are covered [9].

IV. PARALLEL CORPUS BUILDING

For Myanmar NLP researchers, there are many difficulties which are arisen from the lack of resources; in particular parallel corpora are scare [10]. Currently, there is no specific parallel corpus that can be used for Myanmar Primary (MP) students who would like to study Myanmar-English and vice versa translation. Therefore, as a first step, we are building a textual parallel MP corpus with the purpose of developing a Machine Translation (MT)-based approach for using technology to assist for MP students in their educational life.

For this purpose, we collected the necessary data as main sentences from old and new English textbooks (including Grade 1 to 3) which are published by the Ministry of Education. Then, we translated them in Myanmar by using “The Khit Thit English-Myanmar Pocket Dictionary”, compiled by Khit Thit editorial staff and “Pocket Best Speaking” by Professor Minn Nandar (Dr. Min Tin Mon). As in nature of Myanmar language, we found the fact that one English sentence can be translated into many Myanmar sentences. Therefore, we prepared two forms of translation style in our MP corpora such as one-to-one and many translations.

A. One-to-One Translation

In this translation form, one English sentence is translated into one meaningful, polite and written form of Myanmar sentence as follows:

English: My name is Mg Mg.
Myanmar: ကျွန်ုပ် ၏ အမည် သည် မောင်မောင် ဖြစ် ပါ သည် ။

English: I like this cake.
Myanmar: ကျွန်ုပ် သည် ဤ ကိတ်မုန့် အား ကြိုက်နှစ်သက် ပါ သည် ။

B. *One-to-Many Translation*

Unlike the above translation form, one English sentence is translated into many possible Myanmar sentences because there are many types of Myanmar pronouns, daily conversation style and sentence ending words.

For example, "I" can be translated into "ကျွန်ုပ် (I)", "ကျွန်တော် (I)", "ကျနော် (I)", "ကျွန်မ (I)", "ကျမ (I)", "ကျုပ် (I)", "ငါ (I)" and "is" can be "ရှိပါသည် (is)", "ရှိသည် (is)", "ရှိပါတယ် (is)", "ရှိတယ် (is)", "ရှိတယ်လေ (is)", "ဖြစ်ပါသည် (is)", "ဖြစ်သည် (is)", "ဖြစ်ပါတယ် (is)", "ဖြစ်တယ် (is)", "ဖြစ်တယ်လေ (is)" in Myanmar.

For example, we can translate the "My name is Mg Mg." English sentence into several Myanmar sentences as shown below.

English: My name is Mg Mg.
Myanmar: ကျွန်ုပ် ၏ အမည် သည် မောင်မောင် ဖြစ် သည် ။
ကျွန်ုပ် ရဲ့ နာမည် က မောင်မောင် ဖြစ် ပါတယ် ။
ကျွန်ုပ် ရဲ့ နာမည် က မောင်မောင် ဖြစ် တယ် ။
ကျွန်ုပ် ရဲ့ နာမည် က မောင်မောင် ပါ ။
ကျွန်ုပ် ရဲ့ နာမည် က မောင်မောင် လေ ။
ကျွန်တော့် ရဲ့ နာမည် က မောင်မောင် ဖြစ် ပါတယ် ။
ကျွန်တော့် နာမည် က မောင်မောင် လေ ။
ကျုပ် နာမည် က မောင်မောင် ပါ ။
ငါ့ နာမည် က မောင်မောင် လေ ။

In this study, we prepared MP one-to-many corpus that contained 8,394 English sentences and 46,758 translated Myanmar sentences. Some English sentences are translated into 10 or 20 or 30 Myanmar sentences and so on. The maximum number of translations from one English sentence into Myanmar is 1,448.

V. PHRASE-BASED STATISTICAL MACHINE TRANSLATION (PBSMT)

A PBSMT translation model is based on phrasal units [11]. Here, a phrase means a contiguous sequence of words and generally, not a linguistically motivated phrase. Typically, phrase-based translation model gives better translation performance than word-based models. We can describe a simple phrase-based translation model consisting of phrase- pair probabilities extracted from corpus and a basic reordering model, and an algorithm to extract the phrases to build a phrase-table [12].

The phrase translation model is based on noisy channel model. To find best translation $e^{\wedge}$ that maximizes the translation probability P(e|f) given the source sentences; mathematically. Here, the source language is French and the target language is English. The translation of a French sentence into an English sentence is modeled as (1).

$$e^{\wedge} = argmax_{e}\, P(e|f) \tag{1}$$

Applying the Bayes' rule, we can factorized the P(e|f) into three parts as (2).

$$P(e|f) = \frac{P(e)}{P(f)} P(f|e) \tag{2}$$

The final mathematical formulation of phrase-based model is as (3).

$$\text{argmax}_{e}\ P(e|f) = \text{argmax}_{e}\ P(f|e)\ P(e) \tag{3}$$

We note that denominator P(f) can be dropped because for all translations the probability of the source sentence remains the same. The P(e|f) variable can be viewed as the bilingual dictionary with probabilities attached to each entry to the dictionary (phrase table). The P(e) variable governs the grammatically of the translation and we model it using n-gram language model under the PBSMT paradigm.

VI. EXPERIMENTAL SETUP

A. *Corpus Statistics*

For experiments, both one-to-one and many translations corpora contain sentences from Myanmar primary level English textbooks. Myanmar3 Unicode font is used for both Myanmar and English sentences. In one-to-one translation corpus, there are a total of 8,394 parallel sentences: 6,716 (80% of total sentences) sentences for training, 839 (10% of the remaining total sentences) sentences for development and evaluation. And, there are a total of 46,758 parallel sentences in one-to-many translations corpus. As presented in section IV, we considered the fact that one English sentence can be translated into several Myanmar sentences. This makes us more motivation to study the performance of PBSMT with respect to the various Myanmar translated sentences.

Therefore, we prepared different datasets for several experiments. For example, in the dataset for experiment-1, each English sentence is translated into the maximum 10 Myanmar sentences. Similarly, in the dataset for experiment-2, each English sentence is translated into the maximum 20 Myanmar sentences and so on. Therefore, the dataset for final experiment contained the maximum 1,448 Myanmar translated sentences for each English sentence. And, the different datasets are divided into training (80%), development (10%), and testing (10%) datasets respectively. There are no overlap of parallel sentences between training, development, and testing datasets.

B. *Word Segmentation*

A core issue in SMT is the identification of translation units. In phrase-based SMT, these units are comprised of bilingual pairs consisting of sequences of source and target tokens (words). Therefore, word segmentation (which defines the nature of these tokens) is one of the key preprocessing steps in SMT [13]. In this paper, we collect the necessary data as presented in section IV. As we know, Myanmar sentences are written as contiguous sequences of syllables and they are usually not separated by white space. Spaces are used for separating phrases for easier reading. However, it is not strictly necessary, and these spaces are rarely used in short sentences. There are no clear rules for using spaces in Myanmar language, and thus spaces may (or may not) be inserted between words, phrases, and even between a root words and their affixes [13].

In this study, we did manual segmentation process to identify the word boundary by using five rules which are applied by proposed myPOS. These five rules are described with some examples as follows [14]:

- Myanmar word can usually be identified by the combination of root word, prefix and suffix.
  Unsegmented Word: သွားသည်
  Segmented Word: သွား သည်
- Plural Nouns are identified by following the particle.
  Unsegmented Word: ကျောင်းသားများ
  Segmented Word: ကျောင်းသား များ
- Possessive words are identified by following post positional marker.
  Unsegmented Word: သူမ၏ အဖေ
  Segmented Word: သူမ ၏ အဖေ
- Noun is identified with the combination of particle to the verb or the adjective.
  Unsegmented Word: ကျန်းမာရေး၊ ခင်မင်မှု
  Segmented Word: ကျန်းမာ ရေး၊ ခင်မင် မှု
- Particle state the type of noun, and used after number or text number.
  Unsegmented Word: စာအုပ်၂အုပ်၊ ပန်းသီးငါးလုံး
  Segmented Word: စာအုပ် ၂ အုပ်၊ ပန်းသီး ငါး လုံး

Besides, in our manual word segmentation rules, compound nouns are considered as one word and thus, a Myanmar compound word “လက်ဖက်ရည် + အိုး” (“tea” + “pot” in English) is segmented as one word “လက်ဖက်ရည်အိုး”. Myanmar adverb words such as “**စောစောစီးစီး**” (“early” in English) are also considered as one word.

C. *Moses SMT System*

We used the PBSMT system provided by the Moses toolkit [15] for training the PBSMT statistical machine translation systems. The word segmented source language was aligned with the word segmented target language using GIZA++ [16]. The alignment was symmetrized by grow-diag-final and heuristic [11]. The lexicalized reordering model was trained with the msd-bidirectional-fe option [17]. We used KenLM [18] for training the 5-gram language model with modified Kneser-Ney discounting [19]. Minimum error rate training (MERT) [20] was used to tune the decoder parameters and the decoding was done using the Moses decoder (version 2.1.1) [15]. We used default settings of Moses for all experiments.

D. *Evaluation*

Two automatic criteria are used for the evaluation of the machine translation output. One was the de facto standard automatic evaluation metric Bilingual Evaluation Understudy (BLEU) [21] and the other was the Rank-based Intuitive Bilingual Evaluation Measure (RIBES) [22]. The BLEU score measures the precision of n-gram (over all $n \leq 4$ in our case) with respect to a reference translation with a penalty for short translations [21]. Intuitively, the BLEU score measures the adequacy of the translation and large BLEU scores are better. RIBES is an automatic evaluation metric based on rank correlation coefficients modified with precision and special care is paid to word order of the translation results. The RIBES score is suitable for distance language pairs such as Myanmar and English. Large RIBES scores are better.

VII. RESULT AND DISCUSSION

The BLEU and RIBES score results for machine translation experiments with PBSMT between Myanmar and English languages are shown in Table I and II. Here, bold numbers indicate the highest scores among several PBSMT experiments. The RIBES scores are shown in the round brackets. “My” stands for Myanmar, “En” stands for English respectively.

In one-to-one MT, English-Myanmar translation achieved 59.28 BLEU and 0.8468 RIBES scores and Myanmar-English translation achieved 89.42 and 0.9077 RIBES scores using PBSMT approach.

When we measured the performance of PBSMT using one-to-many translation corpora, we found that the BLEU and RIBES scores are gradually increased in both English to Myanmar and Myanmar to English translations as shown in Table II. We carried out these machine translation experiments by incrementing the number of translated Myanmar sentences. In other words, each English sentence is translated into 10, 20, 30, ..., 100, 200, 300, …, 1,448 translated Myanmar sentences.

From the English to Myanmar translation results with one-to-many corpora (see Table II), it can be seen clearly that the gain in BLEU and RIBES scores of 1-20 translation model significantly increased than 1-10 translation model (from 19.68 to 46.28 in terms of BLEU score and from 0.6969 to 0.8118 in terms of RIBES score). From the models 1-30 to 1-90 and 1-200 to 1-1100 translation results, we can assume that increasing the number of translated Myanmar sentences slightly impact (only a small fraction) on PBSMT performance. On the other hand, the results of the 1-100 and 1-1200 models gains significant BLEU scores (+1.64 and -1.52 in average). One more factor we should consider is 839 sentences of the test-set that we used for all one-to-many corpora experiments.

According to the results of Myanmar to English translation models, we found that the gains in BLEU score continuously increased (50.17 ~ 71.15 BLEU) between 1-10 and 1-400 translation model. However, this 71.15 result slightly decreased (average 0.25 BLEU) in 1-500 model. Then, there were alternate changes in terms of BLEU and RIBES scores. The gain in BLEU score of 1-800 translation model is much larger (average 2.18 BLEU) than 1-700 translation model. Generally, English to Myanmar translation results are above 50 BLEU scores for 1-700 to 1-1448 models and the highest BLEU and RIBES scores are achieved by 1-1448 (52.38 BLEU score) and 1-1300 (0.8238 RIBES score) translation models. Similarly, Myanmar to English translation results are above 70 BLEU scores from 1-400 to 1-1448 models and the highest BLEU and RIBES scores (75.12 and 0.8838) are achieved by 1-1300 translation model. Our results with current one-to-one test dataset indicate that Myanmar to English machine translation is better performance (around 23 BLEU and 0.06 RIBES scores higher) than English to Myanmar translation direction.

TABLE I. BLEU AND RIBES SCORES OF PBSMT FOR ONE-TO-ONE TRANSLATION CORPUS BETWEEN MYANMAR AND ENGLISH

Corpus Size	En-My	My-En
training = 6716 development = 839 testing = 839	59.28 (0.8468)	89.42 (0.9077)

TABLE II. BLEU AND RIBES SCORES OF PBSMT FOR ONE-TO-MANY TRANSLATION CORPUS BETWEEN MYANMAR AND ENGLISH

No. of Myanmar Translated Sentences (En-My) Corpus Size [training, dev, test]	En-My	My-En	No. of Myanmar Translated Sentences (En-My) Corpus Size [training, dev, test]	En-My	My-En
1-10 [3591, 449, 449]	19.68 (0.6969)	50.17 (0.8245)	1-400 [22937, 2866, 2867]	49.71 (0.8209)	71.15 (0.8779)
1-20 [6081, 760, 760]	46.28 (0.8118)	57.05 (0.8439)	1-500 [24813, 3102, 3102]	49.80 (0.8174)	70.90 (0.8746)
1-30 [7878, 985, 985]	46.13 (0.8052)	61.52 (0.8639)	1-600 [25417, 3177, 3177]	49.62 (0.8181)	71.79 (0.8790)
1-40 [9255, 1157, 1157]	46.94 (0.81061)	61.97 (0.8650)	1-700 [25831, 3229, 3229]	50.33 (0.8220)	72.79 (0.8834)
1-50 [10313, 1289, 1289]	47.58 (0.8105)	63.22 (0.8670)	1-800 [26269, 3284, 3284]	50.74 (0.8192)	72.38 (0.8812)
1-60 [11098, 1387, 1387]	46.82 (0.8158)	64.48 (0.8691)	1-900 [26718, 3839, 3839]	51.10 (0.8224)	74.56 (0.8830)
1-70 [11777, 1472, 1472]	46.28 (0.8129)	66.86 (0.8758)	1-1000 [27114, 3388, 3389]	50.79 (0.820577)	74.08 (0.878125)
1-80 [12398, 1550, 1550]	46.96 (0.8110)	67.54 (0.8753)	1-1100 [27297, 3412, 3412]	51.85 (0.8225)	72.95 (0.8827)
1-90 [12951, 1619, 1619]	46.78 (0.8115)	67.54 (0.8773)	1-1200 [27434, 3429, 3429]	50.33 (0.8213)	72.87 (0.8799)
1-100 [13470, 1684, 1684]	48.42 (0.8179)	68.54 (0.8778)	1-1300 [27585, 3448, 3448]	51.13 **(0.8238)**	**75.12** **(0.8838)**
1-200 [17698, 2212, 2212]	49.11 (0.8117)	69.43 (0.8758)	1-1400 [27704, 3462, 3462]	52.09 (0.8224)	74.10 (0.8749)
1-300 [20789, 2599, 2599]	50.07 (0.8216)	69.76 (0.8752)	1-1448 [27741, 3468, 3468]	**52.38** (0.8173)	74.49 (0.8777)

From the overall results of Table I and II, both one-to-one and one-to-many models shown that Myanmar to English machine translation achieved better performance than English to Myanmar translation direction. Here, note on corpus size differences (including development and test datasets) among one-to-many models (see Table II). Although, we cannot directly compare between one-to-one and one-to-many model results, we found that the best BLEU and RIBES scores of one-to-many are lower than one-to-one for both My-En and En-My translation directions (BLEU: 52.38 < 59.28, RIBES: 0.8173 < 0.8468 for En-My and BLEU: 74.12 < 89.42, RIBES: 0.8838 < 0.9077 for My-En).

However, the series of BLEU and RIBES scores of the one-to-many models (see Table II) proved that multiple translations of English to Myanmar gradually increased the machine translation performance for both En-My and My-En.

VIII. ERROR ANALYSIS

For both one-to-one and many translation corpora, we analyzed the translated outputs using Word Error Rate (WER) [23]. We also used the SCLITE (score speech recognition system output) program from the NIST scoring toolkit SCTK version 2.4.10 [24] for making dynamic programming based alignments between reference (ref) and hypothesis (hyp) and calculation of WER. The formula for WER can be stated as (4):

$$\text{WER} = \frac{S+D+I}{N} = \frac{S+D+I}{S+D+C} \quad (4)$$

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 ($N = S + D + C$) [23]. It is needed to note that if the number of insertions is very high, the WER can be greater than 100%.

The following examples show WER calculation on the translated outputs of PBSMT approach for Myanmar-English language pair with two types of corpora. The first one is WER calculation for the use of one-to-one Myanmar translation corpus. For example, scoring I, D and S for the translated Myanmar sentence "ကျနော် ကြိုက် သော အစားအစာ က ပေါင်မုန့် ဖြစ် ပါ တယ် ။" ("My favourite food is bread ." in English) compare to a reference sentence, the output of the SCLITE program is as follows:

Scores: (#C #S #D #I) 9 1 1 0
REF: ငါ ရဲ့ ကြိုက် သော အစားအစာ က ပေါင်မုန့် ဖြစ် ပါ တယ် ။
HYP: ********* ကျနော် ကြိုက် သော အစားအစာ က ပေါင်မုန့် ဖြစ် ပါ တယ် ။
Eval: D S

In this case, one substitution (*** ==> ငါ) and one deletion (ရဲ့ ==> ကျနော်) happened, that is C = 9, S = 1, D = 1, I = 0, N = 11 and and thus its WER is equal to 18%. The following is for Myanmar-English translation example and all translated words are correct, C = 6, S = 0, D = 0, I = 0, N = 0 and its WER is equal to 0%.

Scores: (#C #S #D #I) 6 0 0 0
REF: my favourite food is bread .
HYP: my favourite food is bread .
Eval:

The next one is WER calculation for English-Myanmar with one-to-many Myanmar translation corpus. For example, scoring I, D and S for the translated Myanmar sentence "ဒါ က ငါ့ အဘိုး ဖြစ် ပါတယ် ။" ("This is my grandfather.") in English) compare to a reference sentence, the output of the SCLITE program is as follows:

Scores: (#C #S #D #I) 7 1 0 1
REF: ဒါ က ငါ့ *** အဘိုး ဖြစ် ပါတယ် ။
HYP: ဒါ က ငါ့ ရဲ့ *** GRANDFATHER ဖြစ် ပါတယ် ။
Eval: I S

In this case, one substitution (*** ==> ရဲ့) and one insertion (အဘိုး ==> GRANDFATHER) happened, that is C = 7, S = 1, D = 0, I = 1, N = 8 and thus its WER is equal to 25%. The following is for Myanmar-English translation example. In this case, one substitution အဘိုး ==> GRANDFATHER) happened, that is C = 4, S = 1, D = 0, I = 0, N = 5 and thus its WER is equal to 20%.

Scores: (#C #S #D #I) 4 1 0 0
REF: this is my GRANDFATHER .
HYP: this is my အဘိုး .
Eval: S

Fig. 1 and 2 present the average WER percentages of one-to-one and one-to-many translation models. The results show that "Myanmar-English" translation gave the lower WER value than "English-Myanmar" translation.

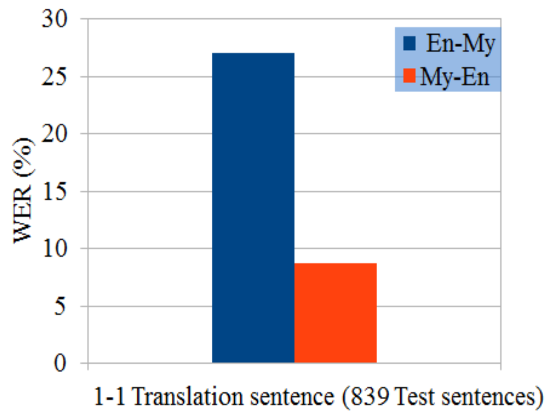

Fig. 1. Average WER% for PBSMT, with one-to-one translation corpus (lower is better)

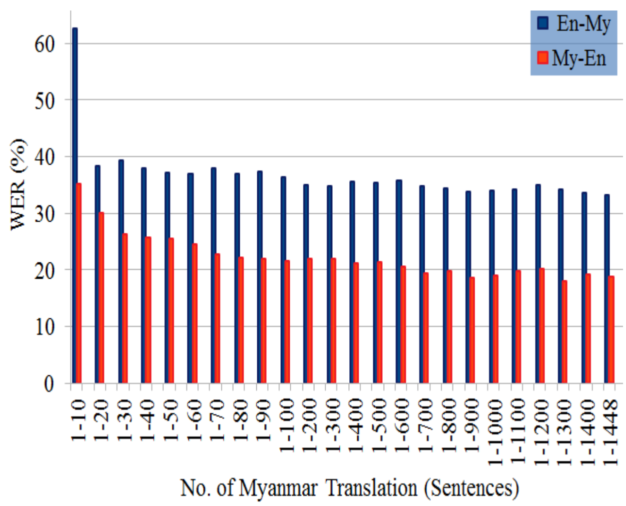

Fig. 2. Average WER% for PBSMT, with one-to-many translation corpus (lower is better)

After we made analysis of the confusion pairs of PBSMT model in details, we found that most of the confusion pairs are caused by (1) the nature of Myanmar languages (written or speaking form), (2) unknown short form (3) ambiguous article mistakes and (4) limited size of the training data especially on English language. For example, the top 10 confusion pairs of one-to-many translation corpus based PBSMT translation model are shown in Table III. In this table, the 1st column is the reference and hypothesis pair (i.e. output of the PBSMT translation model) for English to Myanmar translation. The third one is for that of Myanmar to English translation.

All of the confusion pairs in 1st column are caused by the nature of Myanmar language. For example, in Myanmar written or speaking form, the words "ဟုတ် ("is" in English)" are the same with the words "ဖြစ် ("is" in English)". Also, the words "က ("is" in English)" and "သည် ("is" in English)" in the subject place and the words "ရဲ့ ("of or 's" in English)" and "၏ ("of or 's" in English)" in the possessive place are the same meanings. In other words, these hypotheses are synonyms of the reference words.

TABLE III. THE TOP 10 CONFUSION PAIRS OF PBSMT MODEL USING ONE-TO-MANY TRANSLATION CORPUS BETWEEN MYANMAR AND ENGLISH

En-My (Ref ➔ Hyp)	Freq	My-En (Ref ➔ Hyp)	Freq
ရဲ့ ➔ ၏	29	i'm ➔ am	14
ဟုတ် ➔ ဖြစ်	25	uncle ➔ ဦးလေး	8
ဒီဟာ ➔ ဒါ	24	window ➔ door	8
နေ ➔ ဖြစ်	24	pudding ➔ ပူတင်း	7
အဲ့ဒါ ➔ အဲဒါ	22	am ➔ the	6
ဒါ ➔ ဒီဟာ	21	it's ➔ is	6
ဒီဟာ ➔ က	17	sing ➔ သီချင်းဆို	6
က ➔ သည်	15	a ➔ an	5
တဲ့ ➔ သော	14	an ➔ a	5
ကျနော့် ➔ ရဲ့	13	aunt ➔ အဒေါ်	4

Also, for the machine translation from Myanmar to English, the confusion pairs of "i'm ➔ am" and "it's ➔ is" are caused due to unknown short form. And, we found that the confusion pairs of "am ➔ the", "a ➔ an" and "an ➔ a" are caused by the ambiguous article mistakes. And, the confusion pairs of "window ➔ door" and "pudding ➔ ပူတင်း" and so on are related to the limited size of our training data. Thus, the translation models couldn't learn well.

IX. CONCLUSION

This paper contributes the first PBSMT machine translation evaluation between Myanmar and English languages for specific primary educational domain in Myanmar. We used over 8K Myanmar-English parallel sentences as one-to-one translation corpus and over 46K parallel sentences as one-to-many translation corpus. We analyzed the performance differences of PBSMT translation models by using several number of Myanmar translated sentences (1 English sentence to 10 or 20 of 30 Myanmar translated sentences and so on). The results proved that the highest BLEU and RIBES scores (52.38 and 0.8238 for English-Myanmar and 75.12 and 0.8838 for Myanmar-English) can be achieved for Myanmar-English language pair with one-to-many translation corpus. This paper also presents detail analysis on confusion pairs of machine translation between Myanmar-English and English-Myanmar. In the near future, we plan to extend our experiments with other SMT approaches such as Hierarchical Phrase Based Statistical Machine Translation (HPBSMT) and Operation Sequence Model (OSM) on the one-to-many parallel corpus.

ACKNOWLEDGMENT

We would like to thank Daw Shun Mya Mya, Lecturer at the Department of English, Yangon Technological University for her kind checking to our English to Myanmar translation process.

REFERENCES

[1] Thet Thet Zin, Khin Mar Soe, and Ni Lar Thein, "Myanmar phrases translation model with morphological analysis for statistical Myanmar to English translation system", In Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation, Singapore, December, 2011, pp. 130-139.

[2] Ye Kyaw Thu, Andrew Finch, Win Pa Pa, and Eiichiro Sumita, "A large-scale study of statistical machine translation methods for Myanmar language", In Proceedings of Symposium on Natural Language Processing 2016, February 10-12, 2016.

[3] C. David, Hierarchical Phrase-Based Translation, Computational Linguistics Volume 33, No. (2), June, 2007, pp. 201-228.

[4] P. Kishore, R. Salim, W. Todd, Z. Wei-Jaing, "BLEU: a Method for automatic evaluation of machine translation", IBM Research Report rc22176 (w0109022), Thomas J. Watson Research Center, 2001.

[5] H. Isozaki, T. Hirao, K. Duh, K. Sudoh, and H. Tsukada, "Automatic evaluation of translation quality for distant language pairs", In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, MIT, Massachusetts, USA, October, 2010, pp. 944-952.

[6] Win Pa Pa, Ye Kyaw Thu, Andrew Finch and Eiichiro Sumita, "A study of statistical machine translation methods for under resourced languages", 5th Workshop on Spoken Language Technologies for Under-resourced Languages (SLTU Workshop), Procedia Computer Science, Yogyakarta, Indonesia, Volume 81, May 09-12, 2016, pp. 250–257.

[7] R. Wang, C. Ding, M. Utiyama, and E. Sumita, "English-Myanmar NMT and SMT with Pre-ordering: NICT's Machine Translation Systems at WAT-2018", In Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation: 5th Workshop on Asian Translation, Association for Computational Linguistics, PACLIC Hong Kong, December 01-03, 2018.

[8] New Primary School: https://createmm.org/en

[9] Elementary School Textbooks: https://createmm.org/mm/primary-textbooks

[10] Ye Kyaw Thu, V. Chea, A. Finch, M. Utiyama and E. Sumita, "A Large-scale Study of Statistical Machine Translation Methods for Khmer Language", 29th Pacific Asia Conference on Language, Information and Computation, Shanghai, China, October 30 - November 1, 2015, pp. 259-269.

[11] P. Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation." In Proceedings of HTL-NAACL, 2003, pp. 48–54.

[12] L.Specia, "Tutorial, Fundamental and New Approaches to Statistical Machine Translation", International Conference Recent Advances in Natural Language Processing, 2011.

[13] Ye Kyaw Thu, A. Finch, Y. Sagisaka, and E. Sumita, "A Study of Word Segmentation Schemes for Statistical Machine Translation", In Proceedings of the 11th International Conference on Computer Applications (ICCA 2013), Yangon, Myanmar, February 26-27, 2013, pp. 167-176.

[14] myPOS (Myanmar Part-of-Speech Corpus): https://github.com/ye-kyaw-thu/myPOS

[15] 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.

[16] 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.

[17] T. Christoph, "A Unigram Orientation Model for Statistical Machine Translation", In Proceedings of HLT- AACL 2004: Short Papers, Stroudsburg, PA, USA, 2004, pp. 101-104.

[18] 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.

[19] 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.

[20] 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.

[21] Papineni, K., Roukos, S., Ward, T., Zhu, W., BLEU: a Method for Automatic Evaluation of Machine Translation, IBM Research Report rc22176 (w0109022), Thomas J. Watson Research Center , 2001.

[22] H. Isozaki, T. Hirao, K. Duh, K. Sudoh, and H. Tsukada, "Automatic evaluation of translation quality for distant language pairs", In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 944-952.

[23] Wikipedia of Word Error Rate: https://en.wikipedia.org/wiki/Word_error_rate

[24] (NIST) The National Institute of Stan- dards and Technology. Speech recog- nition scoring toolkit (sctk), version: 2.4.10, 2015.