Unsupervised Neural Machine Translation between Myanmar Sign Language and Myanmar Language

Swe Zin Moe†, Ye Kyaw Thu‡, Hnin Aye Thant†, Nandar Win Min† and Thepchai Supnithi‡

***Abstract*— This paper investigate the utility of unsupervised Neural Machine translation (U-NMT) on low-resource language pairs: Myanmar sign language (MSL) and Myanmar language. Since state-of-the-art neural machine translation (NMT) require large amount of parallel sentences, which we do not have for pairs we consider. We focus primarily on incorporating two different types of monolingual data: translated Myanmar sentences of primary English and myPOS data, only into our Myanmar langauge side. We found that the incorporating monolingual data achieved higher performance than the baseline approach. We prepared four types of training data for U-NMT models and the results clearly show that using the myPOS corpus on incorporating the Myanmar language monolingual data achieved the highest BLEU scores when compared to other training data.**

***Index Terms*—Machine Translation, Neural Machine Translation, Unsupervised Neural Machine Translation, Myanmar sign language, Myanmar language.**

I. Introduction

THERE are about 4.6% of the population are disable and 1.3% of the population are deaf and hearing impairment in Myanmar [1]. There are four schools for the Deaf children in Myanmar; Mary Chapman School for the Deaf Children in Yangon (est. 1904), School for the Deaf children in Mandalay (est. 1964), Immanuel School for the Deaf in Kalay (est. 2005) and School for the Deaf children in Tamwe, Yangon (est. 2014). In Myanmar, based on the information from these schools, only 0.006% of the Deaf have a university level education. This percentage is very low compared to all the population in Myanmar. Most of the Deaf people are suffering substantial exclusion and isolation from social networks for the hearing. Furthermore, unemployment rates in the deaf community are high and most live in poverty. The main reasons are communication problems and widespread lack of awareness of sign language (SL). SL is the primary means of communication for deaf people, although there are not enough SL interpreters and communication systems in Myanmar.

Our purposes was not only to break down the communication barriers between Deaf and general people but also to raise awareness of Deaf culture and importance of sign language. With these purposes, we develop an automatic machine interpreter that can translate Myanmar spoken or written language and MSL. Machine Translation (MT) of MSL would be useful in enabling hearing people who do not know MSL to communicate with Deaf individuals.

Our research contribution is to investigate U-NMT on low-resource language pairs: Myanmar sign language and Myanmar language. We focus primarily on incorporating two different types of Myanmar language monolingual data: translated Myanmar sentences of primary English data and myPOS [2] data. Another contribution is we are developing MSL corpus and we used the current version of the corpus for our experiments. Furthermore, we can make a comparison between NMT and U-NMT for MSL and Myanmar language machine translation.

The structure of the paper is as follows. In the next section, we present a brief review of machine translation systems for text to SL. Section III presents a sketch of MSL and Myanmar language. Section IV presents preparation of the MSL corpus for machine translation experiments. Then, in Section V, we describe the methodologies used in the machine translation experiment. Section VI presents statistical information of the corpus and the experimental settings. The results together with some discussions are presented in Section VII. Section VIII presents the error analysis of translated sentences. Finally in Section IX, we present our conclusions and indicate promising results for future research.

II. MT for Sign Language

MT systems between spoken and sign languages had a start in the late 90s. Strategies used for developing MT system are also used for developing text to sign language MT system including direct MT, template-based MT, transfer-based MT, interlingua-based MT, rule-based MT, knowledge-based MT, example-based MT, syntax-based MT and statistical-based MT. Details of each strategy can be found in several books as follows: Hutchins and Somers, 1992 [3]; Hutchins, 2000 [4]; Nirenburg and Raskin, 2004 [5]. A number of text to sign language translation systems have been carried out around the world, e.g. TESSA system (Bangham & Cox, 2000) [6], weather reports generate system (Angus & Smith, 1999) [7], ViSiCAST Translator (Safar & Marshall, 2000) [8], TEAM Project (Zhao & Kipper, 2000) [9], ZARDOZ system (Veale & Collins, 1998) [10], ASL Workbench (Armond & Speers, 2001)

† The Authors are with the Department of Information Science, University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myanmar.
‡ The Authors are with NECTEC, National Electronics and Computer Technology Center, Thailand
Corresponding Authors : {swezinmoe.1011, yktnlp}@gmail.com
Manuscript received December 21, 2019; revised March 15, 2020.

[11], South African sign language machine translation system (Zijl & Barker, 2003) [12], TGT system-polish text into sign language (Suszczanska & Szmal, 2002) [13], spatial and planning models of ASL classifier predicates for MT and American sign language generation: Multimodal natural language generation (NLG) with multiple linguistic channels (Huenerfauth, 2004, 2005) [14], [15], [16], [17], [18], [19], [20] experiments in sign language machine translation using examples (Morrissey & Way, 2006) [21] and Morpho-syntax base statistical methods for automatic sign language translation (Stein, Bungeroth, & Ney, 2006) [22]. Most of them are English-to-American Sign Language (ASL).

III. MSL and Myanmar Language

MSL like other known Sign Languages (SLs) depends on three basic factors that are used to represent the Manual Features (MFs): hand shape, hand location and orientation. In addition to the MFs, MSL also has Non-Manual Features (NMFs) that are related to head, face, eyes, eyebrows, shoulders and facial expression like puffed checks and mouth pattern movements. Postures or movements of the body, head, eyebrows, eyes, cheeks, and mouth are used in various combinations to show several categories of information, including lexical distinction, grammatical structure, adjectival or adverbial content, and discourse functions [23]. Grammatical structure that is shown through non-manual signs includes questions, negation, relative clauses [24], boundaries between sentences [25], and the argument structure of some verbs [26]. Similar to ASL and British Sign Language (BSL), MSL use non-manual marking for yes/no questions. They are shown through raised eyebrows and a forward head tilt [27], [28], [29]. Figure1 shows an example of MSL sentence "မိသားစု (family)" "ဘယ်လောက် (how many) + NMFs – chin up and raised eyebrows for wh-question". The meaning of the MSL sentence is "မိသားစု မှာ လူ ဘယ် နှစ် ယောက် ရှိ သလဲ ။" in Myanmar language and "How many people are there in your family?" in English respectively.

Sign language is different in Yangon and Mandalay regions with many dialects. To the best of our knowledge, MSL using in the Mary Chapman School for the Deaf Children, Yangon is mainly different with MSL of Mandalay region. This difference gives the difficulty of communicating and dealing between Deaf or hearing disabilities in different cities. A government project was set up in 2010 to establish a national sign language with the aid of the Japanese Federation of the Deaf.

MSL is a full natural language that includes various linguistic structures (e.g., grammars, vocabularies, word order, etc.) distinct from Myanmar written language. Myanmar language is tonal and syllable-based. Examples of different grammar, word order and vocabulary used between Myanmar and MSL can be seen in the followings.

English: What time do you wake up?
Myanmar: ဘယ် အချိန် အိပ်ယာ က ထ သလဲ ။

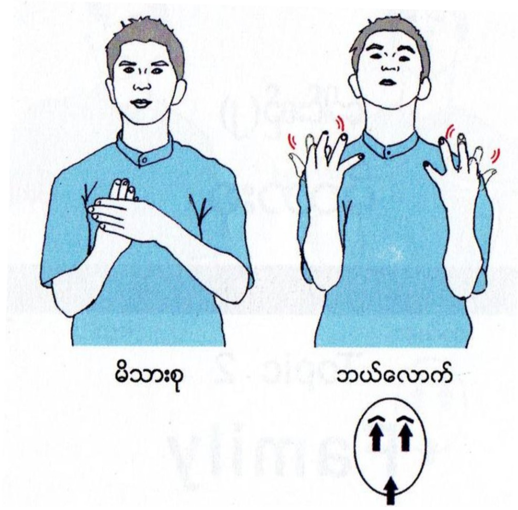

Fig. 1. An example of MSL sentence that used non-manual features [29]

MSL: အိပ်ယာထ (wake up) အချိန် (time) ဘာလဲ (what)

English: I wake up at six o'clock.
Myanmar: မနက် ခြောက် နာရီ မှာ ထ လေ့ ရှိ ပါတယ် ။
MSL: မနက် (morning) နာရီ (o'clock) ခြောက် (six)

English: Daughter-in-law
Myanmar: ချွေးမ ။
MSL: သား (son) လက်ထပ် (marries) မိန်းကလေး (girl)

IV. Corpus Preparation

Myanmar natural language processing (NLP) researchers are facing with many difficulties arising from the lack of resources; in particular parallel corpora are scarce [33]. Currently, there is no parallel corpus for MSL. Therefore, we began building multimedia parallel MSL corpus in October 2016, with the purpose of developing a MT-based approach for using technology to assist hearing and speaking disabilities with limited Myanmar language in their daily life basic conversation.

For this purpose data collection with 30 SL trainers and Deaf people: males and females, age range from 11 to 48, from School for the Deaf (Mandalay), Mary Chapman School for Deaf Children (Yangon), School for the Deaf (Tamwe), Myanmar Deaf Society and Literacy and Language Development for the Deaf in Yangon and Mandalay regions has been carried out. We also considered covering different MSL dialects.

The MSL corpus contains MSL video, a textual representation of Myanmar sign language and translated Myanmar written text. This corpus is beneficial not only to NLP research but also to hearing-impaired and deaf individuals, as it helps them to recognize and respect their language differences and communication styles. To the best of our knowledge, this is the first MSL corpus developed for both academic and public use. Our MSL corpus building is work in progress and MSL video, translated MSL sentences and transcript Myanmar language sentences for Emergency (health, accident, police, fire, earthquake, flood and storm) are publicly available (https://github.com/ye-kyaw-thu/MSL4Emergency).

V. Experimental Methodology

In this section, we describe the methodology used in the unsupervised NMT and NMT experiments for this paper.

A. Unsupervised NMT

Lample et al. (EMNLP 2018) [30] introduce a new unsupervised NMT method, which is derived from earlier work by Artetxe et al. (2018) [31] and Lample et al. (ICLR 2018)[32].

1) Initialization

While prior work relied on bilingual dictionaries, [30] propose a more effective and simpler approach which is particularly suitable for related languages. First, instead of considering words, [30] consider byte-pair encodings (BPE) (Sennrich et al., 2015b) [33], which have two major advantages: they reduce the vocabulary size and they eliminate the presence of unknown words in the output translation. Second, instead of learning an explicit mapping between BPEs in the source and target languages, [30] define BPE tokens by jointly processing both monolingual corpora. If languages are related, they will naturally share a good fraction of BPE tokens, which eliminates the need to infer a bilingual dictionary. In practice, i) join the monolingual corpora, ii) apply BPE tokenization on the resulting corpus, and iii) learn token embeddings (Mikolov et al., 2013) [34] on the same corpus, which are then used to initialize the look up tables in the encoder and decoder.

2) Language Modeling

In NMT, language modeling is accomplished via denoising autoencoding, by minimizing:

$$\begin{aligned} \mathcal{L}^{lm} \quad = \quad & \mathbb{E}_{x\sim\mathcal{S}}[-\log P_{s\to s}(x|C(x))]+ \\ & \mathbb{E}_{y\sim\mathcal{T}}[-\log P_{t\to t}(y|C(y))] \end{aligned} \tag{1}$$

where C is a noise model with some words dropped and swapped as in Lample et al. (ICLR 2018) [32]. $P_{s\to s}$ and $P_{t\to t}$ are the composition of encoder and decoder both operating on the source and target sides, respectively.

3) Back-translation

Let us denote by $u^*(y)$ the sentence in the source language inferred from $y \in \mathcal{T}$ such that $u^*(y) = \text{argmax} P_{t\to s}(u|y)$. Similarly, let us denote by $v^*(x)$ the sentence in the target language inferred from $x \in S$ such that $v^*(x) = \text{argmax} P_{t\to s}(v|x)$. The pairs $(u^*(y), y)$ and $(x, v^*(x)))$ constitute automatically-generated parallel sentences which, following the back-translation principle, can be used to train the two MT models by minimizing the following loss:

$$\begin{aligned} \mathcal{L}^{back} \quad = \quad & \mathbb{E}_{y\sim\mathcal{T}}[-\log P_{s\to t}(y|u^*(y))]+ \\ & \mathbb{E}_{x\sim\mathcal{S}}[-\log P_{t\to s}(x|v^*(x))]. \end{aligned} \tag{2}$$

Note that when minimizing this objective function Lample et al. [30] do not back-prop through the reverse model which generated the data, both for the sake of simplicity and because [30] did not observe improvements when doing so. The objective function minimized at every iteration of stochastic gradient descent, is simply the sum of $\mathcal{L}^{lm}$ in Eq.(1) and $\mathcal{L}^{back}$ in Eq.(2). To prevent the model from cheating by using different subspaces for the language modeling and translation tasks, [30] add an additional constraint which is discuss next.

4) Sharing Latent Representations

A shared encoder representation acts like an interlingua,which is translated in the decoder target language regardless of the input source language. This ensures that the benefits of language modeling, implemented via the denoising autoencoder objective, nicely transfer to translation from noisy sources and eventually help the NMT model to translate more fluently. In order to share the encoder representations, [30] share all encoder parameters (including the embedding matrices since we perform joint tokenization) across the two languages to ensure that the latent representation of the source sentence is robust to the source language. Similarly, [30] share the decoder parameters across the two languages. While sharing the encoder is critical to get the model to work, sharing the decoder simply induces useful regularization. Unlike prior work (Johnson et al., 2016) [35], the first token of the decoder specifies the language the module is operating with, while the encoder does not have any language identifier.

B. Self-attentional Transformer

The transformer model (Vaswani et al., 2017) [36] uses attention to replace recurrent dependencies, making the representation at time step i independent from the other time steps. This requires the explicit encoding of positional information in the sequence by adding fixed or learned positional embeddings to the embedding vectors.

The encoder uses several identical blocks consisting of two core sublayers, self-attention and a feed-forward network. The self-attention mechanism is a variation of the dot-product attention (Luong et al., 2015) [37] generalized to three inputs: a query matrix $\mathbf{Q} \in \mathbb{R}^{n\times d}$, a key matrix $\mathbf{K} \in \mathbb{R}^{n\times d}$, and a value $\mathbf{V} \in \mathbb{R}^{n\times d}$, where d denotes the number of hidden units. [36] further extend attention to multiple *heads*, allowing for focusing on different parts of the input. A single *head* u produces a context matrix

$$\mathbf{C}_u = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{W}_u^Q(\mathbf{K}\mathbf{W}_u^K)^T}{\sqrt{d_u}}\right)\mathbf{V}\mathbf{W}_u^V, \tag{3}$$

where matrices $\mathbf{W}_u^Q$, $\mathbf{W}_u^K$ and $\mathbf{W}_u^V$ are in $\mathbb{R}^{d\times d_u}$. The final context matrix is given by concatenating the heads, followed by a linear transformation: $\mathbf{C} = [\mathbf{C}_1; \ldots; \mathbf{C}_h]\mathbf{W}^O$.

The form in Equation (3) suggests parallel computation across all time steps in a single large matrix multiplication. Given a sequence of hidden states $\mathbf{h}_i$ (or input embeddings), concatenated to $\mathbf{H} \in \mathbb{R}^{n \times d}$, the encoder computes self-attention using $\mathbf{Q} = \mathbf{K} = \mathbf{V} = \mathbf{H}$. The second subnetwork of an encoder block is a feed-forward network with ReLU activation defined as

$$FFN(\mathbf{x}) = \max(0, \mathbf{x}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2, \qquad (4)$$

which is also easily parallelizable across time steps. Each sublayer, self-attention and feedforward network, is followed by a post-processing stack of dropout, layer normalization (Ba et al., 2016) [38], and residual connection.

The decoder uses the same self-attention and feed-forward networks subnetworks. To maintain auto-regressiveness of the model, self-attention on future time steps is masked out accordingly [36]. In addition to self-attention, a source attention layer which uses the encoder hidden states as key and value inputs is added. Given decoder hidden states $\mathbf{S} \in \mathbb{R}^{m \times s}$ and the encoder hidden states of the final encoder layer $\mathbf{H}^l$, source attention is computed as in Equation (3) with $\mathbf{Q} = \mathbf{S}, \mathbf{K} = \mathbf{H}^l, \mathbf{V} = \mathbf{H}^l$. As in the encoder, each sublayer is followed by a post-processing stack of dropout, layer normalization (Ba et al., 2016) [38], and residual connection.

VI. Experiments

The first describe the datasets and experimental protocol we used. The next subsections provide details about the architecture and training procedure of our models.

A. Corpus statistics

The MSL corpus is a collection of everyday basic conversation expressions. It contains six main categories and they are people (greeting, introduction, family, daily activities, education, occupations, and communication), food (food, beverage and restaurant), fun (shopping, hobbies and sports), resource (number, time, weather and accuracy), travel (bus, train and airport) and emergency (health, accident, police, fire, earthquake, flood and storm).

For NMT experiments, (5,740) parallel MSL and Myanmar language sentences of our MSL corpus, 4,500 sentences for training, 650 sentences for development and 590 sentences for evaluation were used. For U-NMT experiments, we used only 4,500 MSL sentences for all U-NMT models, since MSL data are scarce and no more data to incorporate. We prepared four types of Myanmar language monolingual data by incorporating translated Myanmar sentences of primary English and myPOS [2] corpus to our existing scarce MSL data for training U-NMT models. Table I presents the preparation of four types of Myanmar language monolingual data and total number of sentences.

The myPOS Corpus (Myanmar Part-of-Speech Corpus) is a 11,000 sentences (264,920 words or 242,865 words if we consider compound words) manually word segmented and POS tagged corpus developed for Myanmar language NLP research and developments. (Khin War War Htike et al.,) [39] collected Myanmar sentences from Wikipedia that include various area such as economics, history, news, politics and philosophy.

The primary English data is a 15,476 sentences (111,075 words) manually word segmented, collected from Grade 1, 2 and 3 textbooks of the government of the Republic of the Union of Myanmar, Ministry of Eduction [40] and we translated it into Myanmar sentences. This corpus contains two types Myanmar language styles, one is literary or written style and colloquial or spoken style. The differences between written and spoken styles of Myanmar language mostly occur in postpositional marker and particle (Okell et al., 1994 1994) [41]. The followings show the example differences between written and spoken style of Myanmar language ("He is a boy.", in English).

Written style:
သူ ယောက်ျားလေး တစ် ယောက် ဖြစ် ပါ သည် ။
သူ ယောက်ျားလေး တစ် ယောက် ဖြစ် သည် ။
သူ သည် ယောက်ျားလေး တစ် ယောက် ဖြစ် ပါ သည် ။
သူ သည် ယောက်ျားလေး တစ် ယောက် ဖြစ် သည် ။

Spoken style:
သူ ယောက်ျားလေး တစ် ယောက် ပါ ။
သူ ယောက်ျားလေး တစ် ယောက် လေ ။

TABLE I
Four types of Myanmar language monolingual data

	No. Myanmar language sentences	Corpus
Training data 1	4,500	MSL-Myanmar
Training data 2	4,500 15,476 total (19,976)	MSL-Myanmar Primary English
Training data 3	4,500 11,000 total (15,500)	MSL-Myanmar myPOS
Training data 4	4,500 15,476 11,000 total (30,976)	MSL-Myanmar Primary English myPOS

B. Training

This subsection provide details about the architecture and training details of our experiments. All the models are trained with 2 GeForce GTX 1080 8GB ROG STRIX GPUs.

1) NMT

For training NMT models, we used the self-attentional transformer (Vaswani et al., 2017) [36] implementation provided by the Sockeye [42] toolkit, which is based on MXNet [43]. Based on the our previous work, exploring the hyperparameter presentation for MSL Neural MT [44], the initial learning rate is set to 0.0002. If the performance on the validation set has not improved for 8 checkpoints, the learning rate is multiplied by 0.7. We set the early stopping patience to 32 checkpoints. All the neural networks have 8 layers. The size of embeddings and hidden states is 512. The size of training batch was set to 256. We apply layer-normalization and label smoothing (0.1) in all models. We tie the source and target embeddings. The dropout rate of embeddings and Transformer locks is set to (0.1). The attention mechanism in Transformer has 8 heads. We trained NMT models for maximum 1500 epoch using the Adagrad (Duchi et al., 2011) [45] optimizer. The BPE models were trained with a vocabulary size of 4,500.

2) Unsupervised NMT

We use the Transformer (Vaswani et al., 2017) [36] implemented in UnsupervisedMT (Lample et al., EMNLP 2018) [30] [46]. For the Transformer, we use 4 layers both in the encoder and in the decoder. Following Press and Wolf (2016) [47], we share all lookup tables between the encoder and the decoder, and between the source and the target languages. The dimensionality of the embeddings and of the hidden layers is set to 512. We used the Adam optimizer (Kingma and Ba, 2014) [48] with a learning rate of 10^{-4}, $\beta_1 = 0.5$, a batch size of 32 and trained with 4,500 BPE. At decoding time, we generate greedily.

C. Evaluation

We used automatic criteria for the evaluation of the machine translation output. The de facto standard automatic evaluation metric Bilingual Evaluation Understudy (BLEU) [54]. The BLEU score measures the adequacy of the translations language pairs such as Myanmar and English. The higher BLEU scores are better.

TABLE II
BLEU scores of NMT

Src-Trg	epoch 500	epoch 1000	epoch 1500
my-sl	16.04	20.46	**23.92**
sl-my	19.46	22.89	**25.28**

TABLE III
BLEU scores of unsupervised NMT for training data 1

Src-Trg	epoch 500	epoch 1000	epoch 1500
my-sl	**28.99**	28.84	28.74
	(22.40)	(23.75)	**(24.01)**
sl-my	**15.04**	13.10	12.52
	(11.58)	**(12.03)**	(11.75)

TABLE IV
BLEU scores of unsupervised NMT for training data 2

Src-Trg	epoch 500	epoch 1000	epoch 1500
my-sl	27.79	**29.06**	27.69
	(27.04)	**(27.56)**	(27.21)
sl-my	**11.21**	9.76	9.61
	(10.47)	(9.04)	(8.08)

TABLE V
BLEU scores of unsupervised NMT for training data 3

Src-Trg	epoch 500	epoch 1000	epoch 1500
my-sl	28.98	**30.13**	30.04
	(28.30)	(29.40)	**(29.53)**
sl-my	**12.06**	9.88	9.51
	(9.98)	(9.04)	(8.72)

TABLE VI
BLEU scores of unsupervised NMT for training data 4

Src-Trg	epoch 500	epoch 1000	epoch 1500
my-sl	26.76	28.19	**28.88**
	(24.33)	(27.03)	**(27.52)**
sl-my	6.34	7.44	**8.44**
	(5.76)	(6.81)	**(6.88)**

VII. Result and Discussion

Table II shows the BLEU scores of NMT experiments. The BLEU score results of the U-NMT for training data 1, 2, 3 and 4 are presented in Table III, IV, V and VI, respectively. Bold numbers indicate the highest scores among the NMT and U-NMT for four types of training data. We run five U-NMT experiments for each training data-set and shown both highest and average BLEU scores. The average BLEU scores are shown in brackets.

When we focus on U-NMT models, we found that the incorporating monolingual data achieved higher performance when compared to training on the limited bilingual corpus collected for these language pairs. The results clearly shown that training data 3 (using the myPOS corpus on incorporating the Myanmar language monolingual data) achieved the highest BLEU scores among the other training data (see Table V). For the Myanmar to MSL translation, training data 3 gave the highest BLEU 30.13. For the MSL to Myanmar translation, the BLEU score of training data 1 (existing scarce MSL-Myanmar corpus) is 2.98 higher than that of training data 3 (see Table III and V). Training data 2 (incorporating primary English data on the Myanmar language data) gave the higher performance than the training data 4 (incorporating both primary English and myPOS data on the Myanmar language data) (see Table IV and VI).

From the overall results reported in Table II, III, IV, V and VI show that the U-NMT largely outperform NMT on the Myanmar to MSL translation task and no improvements are seen for MSL to Myanmar translation. Though the BLEU score of U-NMT for MSL to Myanmar translation is not as high as state-of-the-art supervised NMT model, it will certainly help us translate in low-resource MSL and Myanmar language pairs.

VIII. Error Analysis

In this paper, we focus on the performances of U-NMT approach. We analyzed the translated outputs of unsupervised NMT models using Word Error Rate (WER). We used SCLITE (score speech recognition system output) program from the NIST scoring toolkit SCTK version 2.4.10 (http://www1.icsi.berkeley. edu/Speech/docs/sctk-1.2/sclite.htm) for making dynamic programming based alignments between reference (ref) and hypothesis (hyp) and calculation of WER. The formula for WER can be stated as equation (5):

$$WER = (I + D + S)100/N \tag{5}$$

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 ($N = S + D + C$) [55]. Note that if the number of insertions is very high, the WER can be greater than 100%.

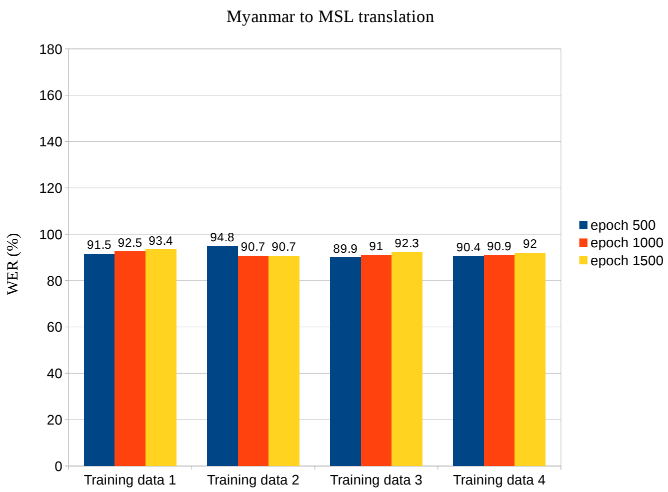

Fig. 2. WER of unsupervised NMT approach for Myanmar to MSL translation

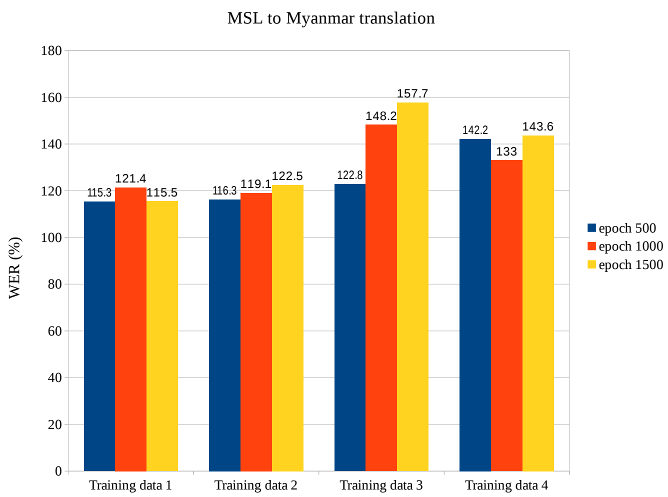

Fig. 3. WER of unsupervised NMT approach for MSL to Myanmar translation

Figure 2 and 3 present the WER percentages of translation between MSL and Myanmar. The results show that

TABLE VII
TOP 10 CONFUSION PAIRS OF UNSUPERVISED NMT MODEL FOR MYANMAR TO MSL TRANSLATION

Ref-Hyp of the unsupervised NMT model				
1:	23	->	◌့	==>	◌်
2:	18	->	ပါ	==>	◌်
3:	13	->	◌်	==>	◌့
4:	12	->	မ	==>	◌်
5:	11	->	မှာ	==>	◌်
6:	11	->	◌့	==>	◌း
7:	10	->	တာ	==>	◌်
8:	9	->	နဲ	==>	◌်
9:	8	->	ပါ	==>	◌း
10:	8	->	◌်	==>	◌း

TABLE VIII
TOP 10 CONFUSION PAIRS OF UNSUPERVISED NMT MODEL FOR MSL TO MYANMAR TRANSLATION

Ref-Hyp of the unsupervised NMT model					
1:	26	->	◌်	==>	◌့
2:	24	->	◌်	==>	ပါ
3:	14	->	◌်	==>	တာ
4:	13	->	◌်	==>	နေ
5:	13	->	◌်	==>	◌း
6:	12	->	◌်	==>	ပဲ
7:	11	->	ပါ	==>	ဘာလဲ
8:	8	->	ဘာလဲ	==>	လား
9:	8	->	◌း	==>	ပါ
10:	8	->	◌်	==>	နဲ

training data 3 gave the lowest WER values for Myanmar to MSL translation and the difference is higher for the MSL to Myanmar translation.

From our detail analysis on confusion pairs of U-NMT models, most of the confusion pairs are caused by the four main reasons and they are (1) word embedding scheme (2) the nature of the sign language and Myanmar language (3) limited size of the training data (4) domain of the corpora. For example, the top 10 confusion pairs of U-NMT model for Myanmar to MSL and MSL to Myanmar translations are shown in Table VII and VIII, respectively. Here, confusion pair number 1 to 10 in Table VII and confusion pair number 1 to 6, 9 and 10 in Table VIII are caused by the word embedding scheme and the nature of Myanmar language. Myanmar language is tonal and syllable based. Generally, Myanmar words are composed of multiple syllables and most of the syllables are composed of more than one character. In Myanmar (Burmese) alphabet, final symbol (asat) “◌်” is used over any of the syllable-final consonants when no stacking takes place, eg. “စပ်” [saʔ] (“of taste - hot”, “mix things”, “join”, etc. in English). This character is also used in combination with other characters to produce a vowel plus tone combination, eg. “ဘယ်” [bɛ̀] (“which” in English). Myanmar sign dot below (aukmyit) “◌့” indicates the creaky tone, eg. မေ့ [mḛ] (“forget” in English) but only used with a consonant final. Another Myanmar sign (Visarga) “◌း” indicates the high tone, eg. ကား [ká] (“car” in English) but cannot be used alone [56], [57]. And thus, the translation model couldn't learn well. The confusion pair number 7 and 8 in Table VIII are caused by the nature of sign language and Myanmar language. The Myanmar language makes prominent usage of particles, which are untranslatable words that are suffixed or prefixed to words to indicate the level of respect, grammatical tense, or mood [58] (eg. “လား”, “ပါ”, “တာ”, etc.) and sign language use non-manual signs, which we do not considered in this paper. We assumed this is also relating to the limited size of our training data and domain of the corpora which we used to incorporate.

IX. Conclusion

This paper investigate the utility of unsupervised Neural Machine translation (U-NMT) on low-resource language pairs: Myanmar sign language (MSL) and Myanmar language. We found that the incorporating monolingual data achieved higher performance than the baseline NMT approach. The results show that using the myPOS corpus on incorporating the Myanmar language monolingual data achieved the highest BLEU scores when compared to the other training data. From the overall results show that the U-NMT largely outperform NMT on the Myanmar to MSL translation task and no improvements are seen for MSL to Myanmar translation. Though the BLEU score of U-NMT for MSL to Myanmar translation is not as high as state-of-the-art supervised NMT model, it will certainly help us machine translation for low-resource nonparallel language pairs.

Acknowledgment

We would like to thank principals, teachers, SL trainers and students of School for the Deaf (Mandalay), Mary Chapman Chapman School for the Deaf Children (Yangon), School for the Deaf, Tamwe (Yangon), Myanmar Deaf Society and Literacy and Language Development for the Deaf for their kind contribution to our research. We would like to thank Honey Htun, Department of Computer Engineering and Information Technology, Yangon Technological University (YTU), Yangon, Myanmar, for sharing her translated Myanmar sentences of Primary English corpus.

References

[1] The population and housing census of Myanmar, 2014
[2] https://github.com/ye-kyaw-thu/myPOS myPOS Corpus (Myanmar Part-of-Speech Corpus) for Myanmar language NLP Research and Developments
[3] Hutchins, W. J., “Early years in machine translation”, John Benjamins Publishing, 2000, USA doi: 10.1075/sihols.97
[4] Hutchins, W. J., & Somers, H. L., “An introduction to machine translation”, Academic Press, 1992, London, ISBN-13: 978-0123628305
[5] Nirenburg, S., & Raskin, V., “Ontological semantics”, The MIT Press, 2004 ISBN: 9780262140867
[6] Bangham, J. A., & Cox, S. J., “Signing for the deaf using virtual humans”, In Proceeding of the Speech and Language Processing for Disabled and Elderly People (Ref. No. 2000/025), IEE Seminar, 2000, London, UK
[7] Angus B., & Smith, G., “English to American sign language machine translation of weather reports”, In Proceeding of the 2^{nd} high desert student conference in linguistics (HDSL2), 1999, Albuquerque, New Mexico, pp. 23–30
[8] Safar, E., & Marshall, I., “The architecture of an English-text-to-sign-language translation system”, In Angelova, G. (Ed.), Recent advances in natural language processing (RANLP), 2000, Tzigov Chark, Bulgaria, pp. 223–228
[9] Zhao, L., & Kipper, K., “A machine translation system from English to American sign language”, In Proceeding of the 4^{th} conference of the association for machine translation in the americas on envisioning machine translation in the information future, 2000, Springer-Verlag, pp. 54– 67
[10] Veale, T., & Collins, B., “The Challenges of Cross-modal Translation: English to sign language translation in the ZARDOZ system”, Machine Translation, 13, 1998, pp. 81–106.
[11] Armond, D., & Speers, L., “Representation of American sign language for machine translation”, Ph.D. Thesis, 2001, Department of linguistics, Georgetown University.
[12] Zijl, L. V., & Barker, D., “South African sign language machine translation system”, In Proceeding of the 2^{nd} international conference on computer graphics, virtual reality, visualisation and interaction in Africa (ACM SIGGRAPH), 2003, Cape Town, South Africa, pp. 49–52
[13] Suszczanska, N., & Szmal, P., “Translating Polish text into sign language in the TGT system”, In Proceeding of the 20^{th} IASTED international multi-conference applied informatics AI, 2002, Innsbruck, Austria, pp. 282–287
[14] Huenerfauth, M. (2004a), “A multi-path architecture for machine translation of English text into American sign language animation”, In Proceeding of the student workshop at the human language technology conference/North American chapter of the association for computational linguistics annual meeting (HLT-NAACL), May 02 - 07, 2004, Boston, MA, USA, pp. 25-30
[15] Huenerfauth, M. (2004b), “Spatial and planning models of ASL classifier predicates for machine translation”, In Proceeding of the 10^{th} international conference on theoretical and methodological issues in machine translation (TMI 2004), Baltimore, MD, USA.
[16] Huenerfauth, M. (2004c), “Spatial representation of classifier predicates for machine translation into American Sign Language”, In Proceeding of the workshop on the representation and processing of signed languages, 4^{th} international conference on language resources and evaluation (LREC 2004), Lisbon, Portugal.
[17] Huenerfauth, M. (2005a), “American Sign Language generation: Multimodal NLG with multiple linguistic channels”, In Proceeding of the student research workshop, the 43^{rd} annual meeting of the association for computational linguistics, Ann Arbor, MI, USA.
[18] Huenerfauth, M. (2005b), “American Sign Language, natural language generation and machine translation”, ACM SIGACCESS Accessibility and Computing (Vol. 81). New York: ACM Press.
[19] Huenerfauth, M. (2005c), “American sign language spatial representations for an accessible user-interface”, In Proceeding of the 3^{rd} international conference on universal access in human-computer interaction, Las Vegas, NV, USA.
[20] Huenerfauth, M. (2005d), “Representing coordination and non-coordination in an American sign language animation”, In Proceeding of the 7^{th} international ACM SIGACCESS conference on computers and accessibility (ASSETS 2005), Baltimore, MD, USA
[21] Morrissey, S. & Way, A., “Experiments in sign language machine translation using examples”, In Proceeding of the IBM CASCON 2006 Dublin symposium, Dublin, Ireland.
[22] Stein, D., Bungeroth, J., & Ney, H., “Morpho-syntax based statistical methods for sign language translation”, In Proceeding of the 11^{th} annual conference of the European association for machine translation, June, 2006, Oslo, Norway.
[23] Wikipedia of Fingerspelling: https://en.wikipedia.org/wiki/Fingerspelling
[24] Boudreault, Patrick; Mayberry, Rachel I., “Grammatical processing in American Sign Language:Age of first-language acquisition effects in relation to syntactic structure”. Language and Cognitive Processes, Volume 21, 2006 – Issue 5, pages 608-635, https://doi.org/10.1080/01690960500139363

[25] Fenlon, Jordan; Denmark, Tanya; Campbell, Ruth; Woll, Bencie, “Seeing sentence boundaries”, Sign Language & Linguistics, 2008, 10 (2), pp. 177 – 200. http://dx.doi.org/10.1075/sll.10.2.06fen

[26] Thompson, RobYin; Emmorey, Karen; Kluender, Robert, “The Relationship between Eye Gaze and Verb Agreement in American Sign Language: An Eye-tracking Study”. Natural Language & Linguistic Theory. 24 (2), 2006, pp. 571–604 doi:10.1007/s11049-005-1829-y.

[27] Baker, Charlotte, and Dennis Cokely, “American Sign Language: A teacher’s resource text on grammar and culture”, 1980, Silver Spring, MD: T.J. Publishers.

[28] Sutton-Spence, Rachel, and Bencie Woll, “The linguistics of British Sign Language”, Cambridge: Cambridge University Press, 1998

[29] “Myanmar Sign Language Basic Conversation Book”, Ministry of Social Welfare, Relief and Resettlement, Department of Social Welfare, Japan International Cooperation Agency, 1st Edition, August 2009, Daw Yu Yu Swe, Department of Social Welfare.

[30] G. Lample, M. Ott, A. Conneau, L. Denoyer, MA. Ranzato, “ Phrase-Based & Neural Unsupervised Machine Translation”, Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)

[31] Mikel Artetxe, Gorka Labaka, Eneko Agirre, and Kyunghyun Cho. 2018, “Unsupervised neural ma-chine translation”, In InternationalConference on Learning Representations(ICLR).

[32] G. Lample, A. Conneau, L. Denoyer, MA. Ranzato, “Unsupervised Machine Translation With Monolingual Data Only”, International Conference on Learning Representations (ICLR), 2018

[33] Rico Sennrich, Barry Haddow, and Alexandra Birch.2015b, “Neural machine translation of rarewords with subword units”, In Proceedings of the 54^{th} Annual Meeting of the Association for Computational Linguistics, pages 1715–1725.

[34] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013, “Distributed representations of words and phrases and their compositionality”, In Advances inneural information processing systems, pages 3111–3119.

[35] M. Johnson, M. Schuster, Q.V. Le, M. Krikun,Y. Wu, Z. Chen, N. Thorat, F. Vi´egas, M. Wat-tenberg, G. Corrado, M. Hughes, and J. Dean.2016, “Google’s multilingual neural machine translation system: Enabling zero-shot translation”, In Transactions of the Association for Computational Linguistics.

[36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, “Attention is all you need”. CoRR, abs/1706.03762, 2017.

[37] Thang Luong, Hieu Pham, and Christopher D. Manning, “Effective approaches to attention-based neural machine translation”. In EMNLP, 2015.

[38] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton, “Layer normalization”. CoRR, abs/1607.06450, 2016.

[39] Khin War War Htike, Ye Kyaw Thu, Zuping Zhang, Win Pa Pa, Yoshinori Sagisaka and Naoto Iwahashi, “Comparison of Six POS Tagging Methods on 10K Sentences Myanmar Language (Burmese) POS Tagged Corpus”, at 18th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2017), April 17 23, 2017, Budapest, Hungary.

[40] Grad 1, 2 and 3 textbooks, The Government of the Republic of the Union of Myanmar Ministry of Education, Basic Education Curriculum and Textbook committee, 2015-2016 Academic year, 2014 October.

[41] Okell, J., Tun, S., Swe, K. M., and University., N. I.(1994 1994). Burmese (Myanmar) : an introduction to the spoken language by John Okell ; with assistance from U Saw Tun and Daw Khin Mya Swe.Northern Illinois University, Center for Southeast Asian Studies DeKalb, Illinois.

[42] Felix Hieber, Tobias Domhan, Michael Denkowski, David Vilar, Artem Sokolov, Ann Clifton, and Matt Post. “Sockeye: A Toolkit for Neural Machine Translation”. ArXiv e-prints, December 2017.

[43] Tianqi Chen, Mu Li, Yutian Li, Min Lin, Naiyan Wang, Minjie Wang, Tianjun Xiao, Bing Xu, Chiyuan Zhang, and Zheng Zhang, “Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems”. CoRR, abs/1512.01274, 2015.

[44] Swe Zin Moe, Ye Kyaw Thu, Hnin Aye Thant and Nandar Win Min, “Neural Machine Translation between Myanmar Sign Language and Myanmar Written Text”, In the second Regional Conference on Optical character recognition and Natural language processing technologies for ASEAN languages 2018 (ONA 2018), December 13-14, 2018, Phnom Penh, Cambodia.

[45] Duchi, John C., Elad Hazan, and Yoram Singer, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization”, Journal of Machine Learning Research, 12:2121–2159, 2011.

[46] https://github.com/facebookresearch/UnsupervisedMT

[47] Ofir Press and Lior Wolf. 2016, “Using the output em-bedding to improve language models”, arXivpreprintarXiv:1608.05859.

[48] Diederik Kingma and Jimmy Ba. 2014, “Adam: Amethod for stochastic optimization”, arXivpreprintarXiv:1412.6980.

[49] Holger Schwenk, “Continuous space translation models for phrase-based statistical machine translation”, In COLING, 2012.

[50] Nal Kalchbrenner and Phil Blunsom, “Recurrent continuous translation models”, In EMNLP, 2013.

[51] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le, “Sequence to sequence learning with neural networks”, In NIPS, 2014.

[52] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, “Neural machine translation by jointly learning to align and translate”. CoRR, abs/1409.0473, 2014.

[53] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N. Dauphin, “Convolutional sequence to sequence learning”. CoRR, abs/1705.03122, 2017.

[54] Papineni, K., Roukos, S., Ward, T., Zhu, W., “Bleu: a Method for Automatic Evaluation of Machine Translation”. IBM Research Report rc22176 (w0109022), 2001, Thomas J. Watson Research Center, In ACL ’02 Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, July 07 - 12, 2002, Philadelphia, Pennsylvania, pp. 311-318

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

[56] Myanmar character notes: https://r12a.github.io/scripts/myanmar/block

[57] Wikipedia of Burmese Alphabet: https://en.wikipedia.org/wiki/Burmese_alphabet

[58] Wikipedia of Burmese language: https://en.wikipedia.org/wiki/Burmese_language

Swe Zin Moe is currently an Assistant Lecturer at Myanmar Institute of Information Technology (MIIT) and also a Ph.D candidate at University of Technology (Yatanarpon Cyber City), UTYCC, Pyin Oo Lwin, Myanmar. Her current doctoral thesis research focuses on machine translation between Myanmar sign language and Myanmar written text. She interested in the general and related problems of natural language processing (NLP) such as machine translation, big data analysis and deep learning.

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 MTU, UCSM, UCSY, UTYCC and YTU.

Hnin Aye Thant is currently working as a Professor and Head of Department of Information Science at the University of Technology (Yatanarpon Cyber City) UTYCC, Pyinoolwin 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 have 14 years teaching experiences in Information Technology specialized in Programming Languages (C, C++, Java & 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 have trained 190 Instructors from ten Technological Universities, twelve Computer Universities & 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 fulfil the students' work-readiness skills.

Nandar Win Min is from Myanmar. She was born on 5th May, 1984. She finished her bachelor degree at Computer University (Taunggyi), Myanmar in 2006. Her Master Degree was finished at Univesity of Computer Studies (Mandalay), Myanmar in 2010 and Ph.D degree was got at University of Technology (Yatanarpon Cyber City) in 2014. Now, she is currently working as a lecturer at Information Science department of University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myanmar.

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