# Enhancing Translation of Myanmar Sign Language by Transfer Learning and Self-Training

**Hlaing Myat Nwe** hlaingmyatnwe@jaist.ac.jp
**Kiyoaki Shirai** kshirai@jaist.ac.jp
**Natthawut Kertkeidkachorn** natt@jaist.ac.jp
Japan Advanced Institute of Science and Technology, Nomi, Ishikawa, Japan

**Thanaruk Theeramunkong** tthanaruk@gmail.com
Sirindhorn International Institute of Technology, Thammasat University, Thailand

**Ye Kyaw Thu** yekyaw.thu@nectec.or.th
**Thepchai Supnithi** thepchai@nectec.or.th
National Electronic & Computer Technology Center (NECTEC), Thailand

**Natsuda Kaothanthong** natsuda@siit.tu.ac.th
Sirindhorn International Institute of Technology, Thammasat University, Thailand

**Abstract**

This paper proposes a method to develop a machine translation (MT) system from Myanmar Sign Language (MSL) to Myanmar Written Language (MWL) and vice versa for the deaf community. Translation of MSL is a difficult task since only a small amount of a parallel corpus between MSL and MWL is available. To address the challenge for MT of its being a low-resource language, transfer learning is applied. An MT model is trained first for a high-resource language pair, American Sign Language (ASL) and English, then it is used as an initial model to train an MT model between MSL and MWL. The mT5 model is used as a base MT model in this transfer learning. Additionally, a self-training technique is applied to generate synthetic translation pairs of MSL and MWL from a large monolingual MWL corpus. Furthermore, since the segmentation of a sentence is required as preprocessing of MT for the Myanmar language, several segmentation schemes are empirically compared. Experiments show that both transfer learning and self-training can enhance the performance of the translation between MSL and MWL compared with a baseline model fine-tuned from a small MSL–MWL parallel corpus only.

## 1 Introduction

In Myanmar, approximately 1.1 M of the population is deaf or has a hearing impairment.[1] Hard-of-hearing people have difficulty comprehending spoken languages because they cannot distinguish sounds. They mostly rely on Myanmar Sign Language (MSL) for communication instead of voice. Since the structure of the grammar, syntax, and lexicon of MSL are different from Myanmar Written Language (MWL), both deaf and hearing people find it rather difficult

[1] https://themimu.info/disabilities-dashboard



*Proceedings of Machine Translation Summit XIX, Vol. 1: Research Track*, pages 111–122
September 4–8, 2023, Macau SAR, China.

to learn. In 2010, the Myanmar government launched a project to establish a standard sign language with the aid of the Japanese Federation of the Deaf (Swe, 2010). This highlights the importance of supporting and promoting MSL to ensure the deaf community has equal access to education and opportunities in Myanmar. Currently, a relatively small number, 0.006% of deaf people have a university education. This percentage is significantly smaller than that for the general population of Myanmar. However, there are few (and limited) assistive technologies available for them. Therefore, deaf people require appropriate ways or tools to enhance communication with hearing people as well as to support their education.

Nowadays, machine translation (MT) plays a role in breaking down language barriers and improving communication between people from various cultures and backgrounds. However, translating low-resource languages is still challenging. One of the solutions to this problem is transfer learning. Transfer learning in MT allows models to use knowledge acquired from other languages to enhance their performance in the target language. It can reduce the costs of the construction of large parallel corpora, enabling the development of high-quality MT systems for low-resource languages. Another technique to tackle the sparseness of the data is semi-supervised learning with self-training, which can construct a parallel corpus automatically.

The goal of this paper is to develop a system to translate MSL to MWL and vice versa. This is a difficult task since the available parallel corpora are very limited. To address the challenge of translating this low-resource language, we propose an approach that combines transfer learning and self-training. Although a few studies have so far been made of the translation of MSL as will be reported in subsection 2.2, there has not been any previous attempt to apply those two techniques to the translation of MSL. We also carry out several experiments to empirically investigate how effective the transfer learning and self-training are.

## 2 Related Work

### 2.1 Machine Translation for Low-Resource Languages Using Transfer Learning

Many researchers have explored the use of transfer learning for MT, particularly in low-resource scenarios. Zoph et al. (2016) prove that transfer learning significantly improves BLEU scores for low-resource languages in neural machine translation (NMT). Their method involves training an MT model for a high-resource language pair (the parent model) and transferring some information from it to an MT model for a low-resource language pair (the child model) by using the parameters of the parent model as the initial parameters of the child model. Experimental results show that the performance of the baseline NMT models is improved by an average of 5.6 BLEU on four low-resource language pairs. Dabre et al. (2017) present how the selection of a parent model influences the performance of child models in transfer learning for NMT. The authors analytically show that the use of a parent model with a source language that is the same or linguistically similar to that of a child model yields the best achievement.

Kocmi and Bojar (2018) propose a simple transfer learning method for NMT under low-resource conditions, where a parent model for a high-resource language pair is first trained and then the training is continued by replacing a training corpus with a low-resource language pair. Unlike the method of (Zoph et al., 2016) where the target language of the parent model is supposed to be the same as that of the child model, any language pairs can be used to train the parent model in their method. The child model performs significantly better than the baseline trained on the parallel corpus of low-resource pairs only, even when unrelated languages with different alphabets are used for training the parent model. The authors claim that it is the first attempt to apply this method to various languages.

Maimaiti et al. (2022) propose a language-independent Hybrid Transfer Learning (HTL) method for improving the quality of the translation in NMT for low-resource languages. They point out that the quality of the translation of NMT for morphologically rich languages tends

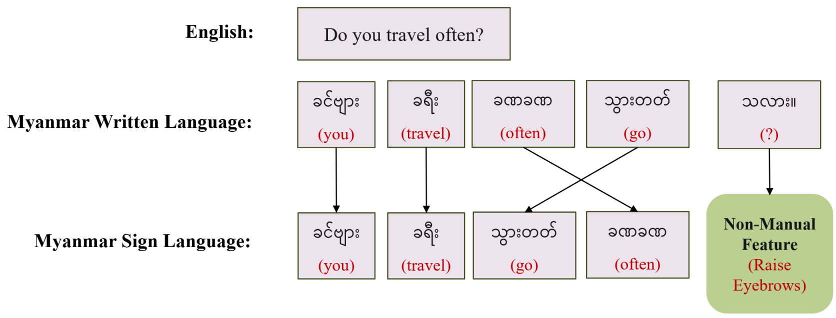


Figure 1: Difference of Grammatical Structure between Myanmar Sign and Written Languages.

to be insufficient due to the sparseness of the data. The suggested HTL approach shares lexicon embeddings between the parent and child languages without using back translation or adding noise manually. According to experimental results, the model trained by the proposed HTL technique consistently exceeds five state-of-the-art methods in the translation of two low-resource languages, namely, Azerbaijani and Uzbek.

### 2.2 Machine Translation for Myanmar Sign Language

Translation of MSL has been a challenging research topic due to the limited amount of available parallel data, which is difficult to construct. Thus, there are few previous studies on MT for MSL. Moe et al. (2018a) evaluate the quality of automatic translation between MSL and MWL using three different statistical machine translation (SMT) approaches and three distinct segmentation schemes and report that Operation Sequence Model and Hierarchical Phrase-based SMT with the syllable-based segmentation achieve the highest performance for translation of MSL $\rightarrow$ MWL and MWL $\rightarrow$ MSL, respectively. The same authors explore NMT approaches and four different segmentation schemes (Moe et al., 2018b). The model based on Transformer (Vaswani et al., 2017) outperforms the Convolutional Neural Network and Recurrent Neural Network in their experiments. They also investigate the utility of unsupervised neural machine translation (U-NMT) on low-resource language pairs, specifically MSL and MWL (Moe et al., 2020). Several monolingual corpora are used and compared for training the NMT model. The highest BLEU score is obtained when the myPOS corpus (Hlaing et al., 2022) is used.

## 3 Myanmar Sign Language

This section briefly introduces the characteristics of MSL, which is the primary communication language for deaf people in Myanmar. To convey meaning, there are two types of features: manual features and non-manual features. The manual features can be categorized into three types: hand shape, hand location, and orientation, which represent words and concepts. To convey additional meanings, MSL also incorporates non-manual features such as movements of the head, eyes, eyebrows, mouth, shoulders, and facial expressions. The facial expressions represent questions, negation, relative clauses, boundaries between sentences, and the argument structure of some verbs. For example, MSL uses non-manual marking, similar to American Sign Language (ASL), to convey yes-or-no questions. That is done by raising the eyebrows and moving the head forward (Boundreault and Mayberry, 2006).

MSL is a natural language with a diverse variety of linguistic features such as grammar, vocabulary, word order, and so on. Such linguistic features are distinct from those of the written language of Myanmar. The Myanmar language is tonal and syllable-based, whereas MSL relies on visual-spatial elements to convey meaning. Additionally, the grammar of MSL and MWL is different. For instance, the grammatical structures of the sentence "Do you travel often?" of MSL and MWL are shown in Figure 1. The word order of MWL is "you," "travel," "often," and "go" followed by the question mark. In contrast, in MSL, the words "often" and "go" are switched reflecting the visual-spatial nature of the sign language. In addition, the question mark is omitted from the word sequence and indicated by a non-manual gesture. That is, they raise their eyebrows to indicate that the sentence is a question.

This study focuses on translating word sequences from MSL to MWL and vice versa. Sentences in MSL are conveyed using glosses, which serve as textual representations of signs. As such, this endeavor can be categorized as a text-to-text translation task, akin to conventional machine translation tasks. Our research may pave the way for a comprehensive system that facilitates seamless conversion between MSL and MWL. However, it's worth noting that while our current approach can convert an MSL gloss into a sign or gesture, it doesn't account for non-manual features – only the manual features represented by words are considered. Expanding the translation process to encompass both manual and non-manual features of MSL remains a challenge for future endeavors.

## 4 Proposed Method

This section describes our proposed method for translation between Myanmar sign and written languages. We use Multilingual Pre-trained Text-to-Text Transfer Transformer (mT5) (Xue et al., 2021) as a base translation system, which is multilingual extension of the Text-to-Text Transfer Transformer (T5) (Raffel et al., 2020). Figure 2 shows a flowchart to train our MT model. Our method consists of two basic methodologies. The first is transfer learning. Firstly, the parent MT model is obtained by fine-tuning the mT5 model using a parallel corpus of high-resource languages, ASL and English. Then the child MT model is trained by fine-tuning the parent model using a relatively small amount of a parallel corpus of the source and target languages. The other basic method is self-training. A new parallel corpus of MSL and MWL is obtained by translating sentences in a monolingual corpus using the initial child MT model. The final MT model is obtained by fine-tuning the parent MT model using the enlarged parallel corpus.

### 4.1 Preprocessing

For preprocessing, sentences in MSL and MWL are segmented into a sequence of tokens. In this study, the following three segmentation schemes are used to split both MSL and MWL sentences and compared in the experiments.

**Word-based Segmentation** A sentence is divided into a word sequence by using spaces. In this study, we manually segment the sentences of MWL using the word-based segmentation rules defined in the previous work (Win et al., 2015). For MSL sentences, segmentation is also manually carried out based on the meaningful MSL word units.

**Syllable-based Segmentation** Myanmar words consist of multiple syllables that usually comprise two or more characters. These syllables are also considered as the basic units for pronouncing Myanmar words. To effectively segment Myanmar syllables, rule-based approaches such as a context-free grammar (Tin, 2012) or regular expressions (RE) can be used. We use the RE-based Myanmar syllable segmentation tool called sylbreak.[2] By syllable-based segmentation, the pronunciation of Myanmar words can be accurately represented and effectively used in the machine learning process.

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

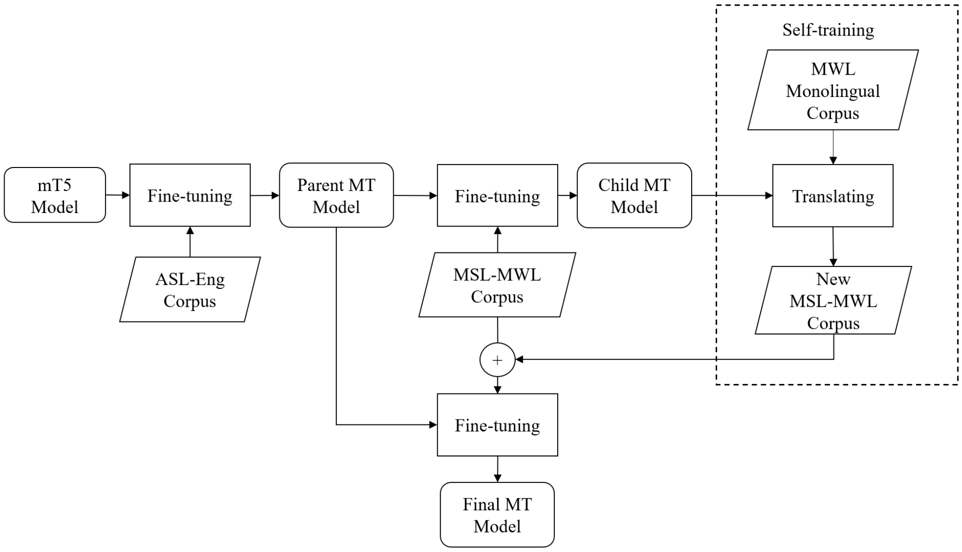


Figure 2: Overview of Proposed Method.

**Byte-pair Encoding (BPE)-based Segmentation** BPE is a method to segment a sentence into subword units. It is particularly effective for handling out-of-vocabulary words, which are words that are not present in the training data. Since BPE builds up a subword vocabulary by merging the most frequently occurring characters, it can handle rare or unknown words by representing them as a combination of common subword units (Sennrich et al., 2016). This study builds the BPE model for segmentation using the subword Neural Machine Translation (subword-nmt)[3] library from the large monolingual corpus of MWL, myPOS corpus, of which the details are described in 5.1.3.

Although the mT5 model has its own tokenizer, in our method, the sentences are split into words, syllables, or BPE by a space and fed into the mT5 model. They are re-tokenized by the mT5 model.

### 4.2 Transfer Learning

Two methods of transfer learning are applied. The first one is to transfer the knowledge obtained from the general pre-trained language model to the task-specific model. Specifically, we use the pre-trained mT5 model, which is applicable to any text-to-text task, for various languages. Among the available five pre-trained mT5 models with different sizes, we choose mT5-Base, which has 580 million parameters.[4] This model is fine-tuned for MT using several parallel corpora.

Another method of transfer learning is a two-step fine-tuning of the mT5 model. First, the parallel corpus of ASL and English is used for fine-tuning the parent MT model. Although the source and target languages are not the Myanmar languages, a relatively large amount of parallel corpus is available. Furthermore, it is supposed that the characteristics of translating between ASL and English and that between MSL and MWL are similar. In other words, the parent MT model can capture some general knowledge about the translation between sign and written languages. Next, the parent MT model is fine-tuned again using a parallel corpus of

[3]https://github.com/rsennrich/subword-nmt

[4]https://github.com/google-research/multilingual-t5

Table 1: Statistics of the Datasets.

| | Parallel | | | | Mono | Parallel* |
|---|---|---|---|---|---|---|
| | ASL–Eng | | MSL–MWL | | MWL | MSL–MWL |
| | Training | Test | Training | Test | | |
| Sentence | 85,710 | 2,000 | 2,836 | 300 | 43,196 | 10,000 |
| Word | 2,131,033 | 50,072 | 36,164 | 3,999 | 537,272 | 92,336 |
| Character | 11,828,933 | 278,499 | 472,044 | 51,905 | 7,534,916 | 1,085,076 |

* automatically constructed by self-training.

MSL and MWL to obtain the child MT model. The knowledge in the parent MT model is transferred to the child MT model, which can compensate for the sparseness of the data of any Myanmar parallel corpus.

### 4.3 Self-Training

A semi-supervised learning approach is applied to improve the MT model between MSL and MWL. We suppose that a small amount of an initial parallel corpus and a large amount of monolingual MWL corpus is available. First, the MT model is trained by transfer learning using the initial parallel corpus as well as the ASL-English parallel corpus. Then, the sentences in the monolingual corpus are translated into MSL sentences using the trained MT model. The pairs of the original and translated sentences form a new parallel corpus. Although it is common to synthesize parallel sentences by back-translation from a parallel corpus, our method generates new samples from a monolingual corpus.

The translated sentences are not always correct, especially when the original sentence is long. To improve the quality of the automatically constructed parallel corpus, unreliable translations are filtered out. To do this, the score of the translated sentence $s$ is calculated using Equation (1),

$$score(s) = \log P(s) \simeq \sum_{w_i \in s} \log P_{mT5}(w_i), \quad (1)$$

where $w_i$ is the $i$-th token in $s$ and $P(s)$ is the probability of generating $s$, while $P_{mT5}(w_i)$ is the probability of generating $w_i$ estimated by the fine-tuned mT5 model. Specifically, at each generation step in the decoder, the distribution of the logits of the mT5 model for all tokens in the vocabulary is converted to the probabilistic distribution by the softmax function. The top $N$ translations with the highest scores are kept to make the new parallel corpus.[5]

## 5 Experiment

### 5.1 Dataset

Three datasets or corpora are used for the experiment. The number of sentences, words and characters of the datasets are summarized in Table 1.

#### 5.1.1 Parallel Corpus of English

The English–ASL Gloss Parallel Corpus 2012 (ASLG-PC12) has been used for training the parent MT model. Due to the absence of a large parallel corpus of sign and written languages in this field, Othman and Tmar (2013) proposed a novel rule-based approach that transformed English part-of-speech (POS) tagged sentences into ASL glosses. This ASLG-PC12 project provided a large parallel corpus consisting of more than one hundred million pairs of sentences

[5]Note that shorter sentences tend to have higher scores and are more likely chosen.

between English and ASL. It includes both manual and non-manual features, where non-manual features are represented by special tokens. The aslg_pc12 dataset[6], which is a part of ASLG-PC12, is used in this experiment. For training the parent model, 85,710 sentences were used as the training data, and 2,000 sentences were used as the test data. Note that the size of the parallel corpus is much larger than the parallel corpus of MSL and MWL reported in 5.1.2.

#### 5.1.2 Parallel Corpus of Myanmar Language

There is only one parallel corpus of MSL and MWL, which was collected from 30 sign language trainers and deaf people. There are 3,136 parallel sentences, from basic conversations in daily life. For our experiment, 2,836 sentences are used for training and 300 for evaluation.

#### 5.1.3 Monolingual Corpus of Myanmar Language

The myPOS corpus[7] was used for self-training. This corpus, also known as the Myanmar POS Tag Corpus, consists of 43,196 sentences that have been manually word-segmented and POS-tagged for the purpose of NLP research and development (Hlaing et al., 2022). The initial child MT model with the word-based segmentation scheme was used to translate the sentences in the myPOS corpus to MSL. In this experiment, the 10,000 sentences that have the highest scores are selected. When training the MT model with the syllable-based and BPE-based segmentation strategies, the sentences in the newly constructed parallel corpus, which are segmented by words, are automatically segmented again by the same strategy.

### 5.2 Experimental Setup

The MT models for both directions, i.e., the models translating from MSL to MWL as well as from MWL to MSL, are trained. Furthermore, several MT models are trained and compared. First, three segmentation schemes (word, syllable, BPE) are used. Second, the models trained with and without transfer learning are compared. Third, the models trained with and without the enlarged parallel corpus obtained by self-training are evaluated.

Two evaluation criteria are used. One is the Bilingual Evaluation Understudy (BLEU) score (Papineni et al., 2002). The bleukit-NTCIR7 Scoring tools[8] is used to calculate the BLEU score. Here, BLEU is measured by counting the overlap of character n-gram to compare MT systems using different segmentation schemes. That is, regardless of the segmentation schemes, the hypothesis and reference sentences are treated as character sequences when BLEU is measured. The other is the Word Error Rate (WER), which is defined by

$$WER = \frac{S + D + I}{N} \tag{2}$$

where $S$, $D$, and $I$ are the number of substitution, deletion, and insertion errors calculated by the alignment between hypothesis and reference sentences, while $N$ is the total number of tokens in the reference. We used the SCLITE[9] (Score Lite) program to get WER.

As already described, the mT5 Base model was utilized as the pre-trained language model. During its fine-tuning, the batch size was set to 20 sentences, and the maximum sequence length was set to 96 tokens so as to handle reasonably long texts. We chose eight hidden layers and six head attention layers, with a hidden layer size of 512. The dropout rate was set to 0.1. The training epochs for the child MT models were set to 500. The number of epochs for training the parent MT model was 10. Servers with NVIDIA A40 and A100 GPUs were used for this experiment.

[6]https://huggingface.co/datasets/aslg

[7]https://github.com/ye-kyaw-thu/myPOS

[8]http://www.nlp.mibel.cs.tsukuba.ac.jp/bleu_kit/

[9]https://github.com/usnistgov/SCTK

Table 2: BLEU Scores and WER of MT Models.

(a) BLEU score (↑)

| Model | MSL→MWL | | | MWL→MSL | | |
|---|---|---|---|---|---|---|
| | word | syllable | BPE | word | syllable | BPE |
| mT5 | 47.77 | 50.67 | 46.30 | 52.79 | 51.23 | 49.62 |
| | [43.95,50.70] | [46.14,54.06] | [43.26,50.11] | [48.80,55.94] | [47.51,54.44] | [45.56,53.00] |
| mT5+T | 49.62 | 51.29 | 46.42 | 52.01 | 56.29 | 50.73 |
| | [45.83,52.58] | [41.89,54.77] | [43.29,49.11] | [47.46,55.22] | [51.80,59.03] | [40.77,54.58] |
| mT5+S | 50.19 | 52.26 | 48.00 | 49.40 | 55.93 | 49.61 |
| | [45.99,54.27] | [48.60,55.89] | [44.47,50.25] | [45.96,52.54] | [52.31,59.37] | [45.94,52.96] |
| mT5+T+S | 51.65 | **56.60** | 53.48 | 56.53 | **57.11** | 51.02 |
| | [47.71,55.29] | [52.72,59.76] | [49.98,56.91] | [52.92,59.84] | [52.61,60.62] | [47.79,52.54] |

(b) WER(%) (↓)

| Model | MSL→MWL | | | MWL→MSL | | |
|---|---|---|---|---|---|---|
| | word | syllable | BPE | word | syllable | BPE |
| mT5 | 53.5 | 50.3 | 52.8 | 57.4 | 51.8 | 51.2 |
| mT5+T | 53.1 | 49.2 | 51.6 | 56.5 | 48.3 | 52.6 |
| mT5+S | 53.9 | 49.7 | 51.2 | 55.9 | 47.9 | 50.8 |
| mT5+T+S | 51.1 | **48.2** | 50.4 | 55.2 | **46.5** | 49.2 |

### 5.3 Results and Discussion

Table 2 (a) shows the BLEU scores with confidence interval values at the significant level of 0.95 of the different MT models. The suffix “+T” in the model name indicates that the MT model is trained by transfer learning with the ASL–English parent MT model. The suffix “+S” indicates that self-training is applied to enlarge the parallel corpus of MSL and MWL. Boldface indicates the best result among the 4 models $\times$ 3 segmentation schemes = 12 MT models.

Among the three segmentation schemes, syllable-based segmentation performs better than the others. The syllable, which represents the pronunciation of a word, might be an appropriate linguistic unit for the translation between MSL and MWL.

Comparing the models mT5 and mT5+T, the use of the parent MT model can improve the BLEU score in most cases. An improvement of 0.62 points in MSL $\rightarrow$ MWL and 5.06 points in MWL $\rightarrow$ MSL with the syllable-based segmentation is found. Transfer learning using the ASL–English parallel corpus is especially effective for translating from written to sign languages. In addition, the quality of the parent MT model has been evaluated. The BLEU scores of the translation of ASL $\rightarrow$ English and English $\rightarrow$ ASL are 85.46 and 98.20 respectively, which are sufficiently high for transfer learning.

Comparing models mT5 and mT5+S, self-training can also boost the BLEU score. The maximum improvement is 4.7 points of the MT model for MWL $\rightarrow$ MSL with the syllable-based segmentation. Self-training is more effective than transfer learning for the translation from MSL to MWL since the BLEU score of mT5+S is better than mT5+T. As for the translation from MWL to MSL, however, transfer learning can improve the performance more as mT5+T is better than mT5+S. Anyway, the contributions of transfer learning and self-training seem comparable, since no significant difference is found between the BLEU scores of mT5+T and mT5+S.

Combining transfer learning and self-training can further boost MT performance since the model mT5+T+S achieves the best BLEU score for all segmentation schemes and translation directions. The highest BLEU scores are 56.60 and 57.11 for MSL $\rightarrow$ MWL and MWL $\rightarrow$ MSL, which are 5.93 and 5.88 points higher than the baseline (the model mT5).

Table 3: Rough Comparison of BLEU Score Between Previous Work and This Study.

| Method | | MSL→MWL | MWL→MSL |
|---|---|---|---|
| mT5+T+S | syllable | 37.83 | 39.97 |
| (Moe et al., 2018a) | Supervised, SMT | 34.78 | 35.11 |
| (Moe et al., 2018b) | Supervised, NMT | 38.21 | 32.92 |
| (Moe et al., 2020) | Unsupervised, NMT | 10.47 | 29.53 |

Table 2 (b) shows the WER of the MT models. The lowest WER is obtained by mT5+T+S with the syllable segmentation scheme. However, since nearly half of the words in translated sentences are errors, there is much room to improve the translation quality. As for the comparison of the models, the results of WER are similar to BLEU, that is, (1) the syllable segmentation is the best, (2) both transfer learning and self-training are effective, and (3) the contributions of those two techniques are comparable.

Table 3 shows the best BLEU scores reported in the previous papers for comparison with our model (mT5+T+S). BLEU score of our method is measured between the hypothesis and reference sentences that are sequences of not characters but syllables, since the previous papers mostly achieved the best results using the syllable segmentation scheme. It is confirmed that the performance of our method is better than or comparable to three previous studies. Note that it is not a fair comparison since the datasets used for the evaluation are different.

### 5.4 Error Analysis

We investigate the errors of the model mT5+T+S for translating from MSL to MWL with the syllable segmentation scheme. In the calculation of WER, three types of errors are considered: a substitution error $S$ (tokens in the reference and output of the MT model are different), a deletion error $D$ (a token in the reference is omitted in the output) and an insertion error $I$ (an extra token is added to the output). The ratios of these errors to the total number of the tokens in the reference are shown in Table 4.

Table 4: Word Error Ratio of Each Type of Error

| $S$ (Substitution) | $D$ (Deletion) | $I$ (Insertion) |
|---|---|---|
| 20.5% | 22.8% | 4.9% |

The most frequent error is a deletion error. This indicates that the word order or grammatical structure is wrong. Example E1 in Figure 3 shows an example of deletion and insertion errors, as well as substitution errors. However, for the purpose of this discussion, we will primarily focus on the deletion and insertion errors. The word "they" is generated as the first word, even though it is the sixth word in the reference. The translation of this example highlights the inability of the model to capture the difference in the grammatical structure between MSL and MWL.

Substitution errors are also often found. This means that the word order is correct, but the word selection is inappropriate. In Example E2, the word "she" in the reference is replaced with "he," causing an inconsistency in the gender with "girl." Besides, some of the substitution errors are not problematic. In Example E3, the Myanmar word "you$_1$" is replaced with the other word "you$_2$." Both words have almost the same meaning but are used in different situations. Specifically, "you$_1$" is used in a business conversation and is never used to talk with family, whereas "you$_2$" is an informal word. Thus, the output is acceptable, although it is different from the reference.

| | | |
|---|---|---|
| E1: | Input: | သူ တို့ (they) နောက် (next) ရာ သီ (weather) နွေ (summer) လက် ထပ် (marry) ။ (end word) |
| | Output: | **သူ တို့** (they) က (preposition) နွေ (summer) ရာ သီ (weather) မှာ (PPM-TIME$_1$) လက် ထပ် (marry) တယ် (PPM-FUTURE) ။ (end word) |
| | Reference: | နောက် (next) လာ မယ့် (coming) နွေ (summer) ရာ သီ (weather) ဆို (PPM-TIME$_2$) **သူ တို့** (they) လက် ထပ် (marry) တော့ မှာ (PPM-PAST) ။ (end word) |
| | English: | They will get married next summer. |
| E2: | Input: | ဟင့် အင်း (no) အ ပျို (single girl) ။ (end word) |
| | Output: | ဟင့် အင်း (no) **ကျွန် တော်** (he) အ ပျို (single girl) ပါ (am) ။ (end word) |
| | Reference: | ဟင့် အင်း (no) **ကျွန် မ** (she) အ ပျို (single girl) ပါ (am) ။ (end word) |
| | English: | No, I am a single girl. |
| E3: | Input: | ခင် ဗျား (you$_1$) လက် (hand) လှုပ် (move) ရ (can) လား (?) ။ (end word) |
| | Output: | **မင်း** (you$_2$) လက် (hand) လှုပ် (move) တတ် (can) လား (can ?) ။ (end word) |
| | Reference: | **ခင် ဗျား** (you$_1$) လက်(hand) လှုပ် (move) လို့ ရ (can) သေး (still) လား (?) ။ (end word) |
| | English: | Can you move your hand? |

PPM: Post-Positional Marker

Figure 3: Example of Errors in Translation from MSL to MWL.

## 6 Conclusion

This paper proposed a novel method to train a machine translation (MT) model for translating between Myanmar Sign Language (MSL) and Myanmar Written Language (MWL). To tackle the problem posed by the fact that MSL is an extremely low-resource language, an mT5 pre-trained model was used as the backbone, and then transfer learning and self-training were applied to improve the quality of the MT system. The contribution of this paper is summarized as follows.

- Transfer learning was first applied for the translation between MSL and MWL. The data of the high-resource language, i.e., the parallel corpus of American Sign Language (ASL) and English, was used to train the parent MT model, and then the knowledge in it was transferred to the child MT model for MSL and MWL.

- Self-training was additionally used to extend the parallel corpus of MSL and MWL that was used for training the child MT model.

- Via the experiments, it was empirically confirmed that both transfer learning and self-training contributed to improving the translation in both directions (MSL $\rightarrow$ MWL and MWL $\rightarrow$ MSL).

In the near future, we will extend our method for the translation of MSL to include both manual and non-manual features. We will also evaluate our MT model from the practical point of view when it is applied for downstream tasks such as cross-lingual information extraction.

## References


Boundreault, P. and Mayberry, R. I. (2006). Grammatical processing in American Sign Language: Age of first-language acquisition effects in relation to syntactic structure. *Language and Cognitive Processes*, 21(5):608–635.

Dabre, R., Nakagawa, T., and Kazawa, H. (2017). An empirical study of language relatedness for transfer learning in neural machine translation. In *Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation*, pages 282–286. The National University (Philipines).

Hlaing, Z. Z., Thu, Y. K., Supnithi, T., and Netisopakul, P. (2022). Improving neural machine translation with POS-tag features for low-resource language pairs. *Heliyon Journal*, 8(8).

Kocmi, T. and Bojar, O. (2018). Trivial transfer learning for low-resource neural machine translation. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 244–252, Brussels, Belgium.

Maimaiti, M., Liu, Y., Luan, H., and Sun, M. (2022). Enriching the transfer learning with pre-trained lexicon embedding for low-resource neural machine translation. *Tsinghua Science and Technology*, 27(1):150–163.

Moe, S. Z., Thu, Y. K., Hlaing, H. W. W., Nwe, H. M., Aung, N. H., Thant, H. A., and Min, N. W. (2018a). Statistical machine translation between Myanmar sign language and Myanmar written text. In *16th International Conference on Computer Applications*, Yangon, Myanmar.

Moe, S. Z., Thu, Y. K., Thant, H. A., and Min, N. W. (2018b). 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)*, Phnom Penh, Cambodia.

Moe, S. Z., Thu, Y. K., Thant, H. A., Min, N. W., and Supnithi, T. (2020). Unsupervised neural machine translation between Myanmar sign language and Myanmar language. *Journal of Intelligent Informatics and Smart Technology*, 1(1):53–61.

Othman, A. and Tmar, Z. (2013). English–ASL gloss parallel corpus 2012: ASLG-PC12, the second release. In *Fourth International Conference On Information and Communication Technology and Accessibility ICTA13*, Hammamet, Tunisia.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, PA, USA.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Sennrich, R., Haddow, B., and Birch, A. (2016). Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany.

Swe, D. Y. Y. (2010). *Myanmar Sign Language Basic Conversation Book*. Ministry of Social Welfare, Relief and Resettlement, Department of Social Welfare, Japan International Cooperation Agency.

Tin, H. H. (2012). Manually constructed context-free grammar for Myanmar syllable structure. In *Proceedings of the Student Research Workshop at the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL12), Association for Computational Linguistics*, pages 32–37, Stroudsburg, PA, USA.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the Advances in Neural Information Processing Systems*, page 5998–6008, Long Beach, CA, USA.

Win, P. P., Ye, K. T., Finch, A., and Sumita, E. (2015). Word boundary identification for Myanmar text using conditional random fields. In *Proceedings of the Ninth International Conference on Genetic and Evolutionary Computing*, pages 447–456, Long Beach, CA, USA.

Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., and Raffel, C. (2021). mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498. Association for Computational Linguistics.

Zoph, B., Yuret, D., May, J., and Knight, K. (2016). Transfer learning for low-resource neural machine translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1568–1575, Austin, TX, USA. Association for Computational Linguistics.