# Machine Translation between Myanmar Sign Language and Myanmar Written Text


Presented by : Swe Zin Moe
2Ph.D(IT)-9

Supervisor : Dr. Hnin Aye Thant
Co Supervisors : Dr. Ye Kyaw Thu, Dr. Nandar Win Min

# Outline

# Abstract


- The proposed system explore Statistical Machine Translation (SMT), Neural Machine Translation (NMT) and Unsupervised NMT (U-NMT) between Myanmar Sign Language (MSL) and Myanmar Written Text (MWT).
- Developing MSL-MWT parallel corpus was used for experiments.
- The experiments were carried out using **three** different **SMT approaches** (phrase-based, hierarchical phrase-based, and the operation sequence model), **three** different **NMT approaches** (Recurrent Neural Network, Transformer, and the Convolutional Neural Network ) and **Unsupervised NMT**.
- Four different segmentation schemes for word embedding were studies, these were syllable segmentation, word segmentation (sign unit based word segmentation for MSL), SentencePiece and the Byte-Pair-Encoding (BPE).
- Highest quality NMT and SMT performances were attained with syllable segmentation for both MSL and MWT.
- From the overall results show that the U-NMT largely outperform NMT on the Myanmar to MSL translation task.

# Introduction

- There are about 4.6% of the population are disable and 1.3% of the population are deaf and hearing impairment in Myanmar.
- 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.
- 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.

# Introduction (con't)

- Sign language is the primary means of communication for deaf people, although there are not enough sign language interpreters and communication systems in Myanmar.
- The purposes of the proposed system are 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, the proposed system develop an automatic machine interpreter that can translate Myanmar spoken or written language and MSL.
- MT of MSL would be useful in enabling hearing people who do not know MSL to communicate with Deaf individuals.
- The challenge for sign language MT is the fact that there is no formal written format for signed languages.
- There are notations systems but no writing system has been adopted widely enough, by the international Deaf community, that it could be considered the 'written form' of a given sign language.
- Sign Languages then are recorded in various video formats. There is no gold standard parallel corpus that is large enough for SMT, for example.

# Contribution

- The proposed system contributes the first evaluation of the quality of automatic translation between Myanmar sign language (MSL) and Myanmar written text (MWT), in both directions.

- The research contribution of this proposed system is to explore SMT, NMT and Unsupervised NMT between MSL and MWT and presenting appropriate hyper-parameters (batch, optimizer) for MWT-MSL and MSL-MWT translations for NMT experiments.

- Another contribution is developing a parallel corpus of MSL and MWT.

- Moreover, the proposed system studied deeply on latest approach Unsupervised NMT (U-NMT) for under-resourced languages and make a comparison among SMT, NMT and (U-NMT).

# Objectives

- To develop an automatic machine interpreter for MSL and Myanmar language
- To built a MSL-MWT parallel corpus
- To investigate the SMT, NMT and unsupervised NMT performances for Myanmar Sign Language (MSL) and Myanmar Written Text (MWT)
- To study the four different word embedding schemes (syllable, word, Byte-Pair-Encoding and SentencePiece) for MSL and MWT translations
- To analyze the translated output sentences of SMT, NMT and U-NMT models
- To find appropriate NMT hyper-parameters for MWT and MSL translation tasks
- To apply Unsupervised NMT approach for under-resource language like MSL and Myanmar language
- To compare the SMT, NMT and Unsupervised NMT performances

# Proposed System Design

Consists of two main components

- Automatic Video Annotation
  - aims for automatic recognition of sign language sentences
  - translate the input video to gloss, a written form of MSL
- Machine Translation
  - translates from sequences of sign identifiers (provided by the automatic video annotation) into text, Myanmar written text

# Proposed System Design (con’t)

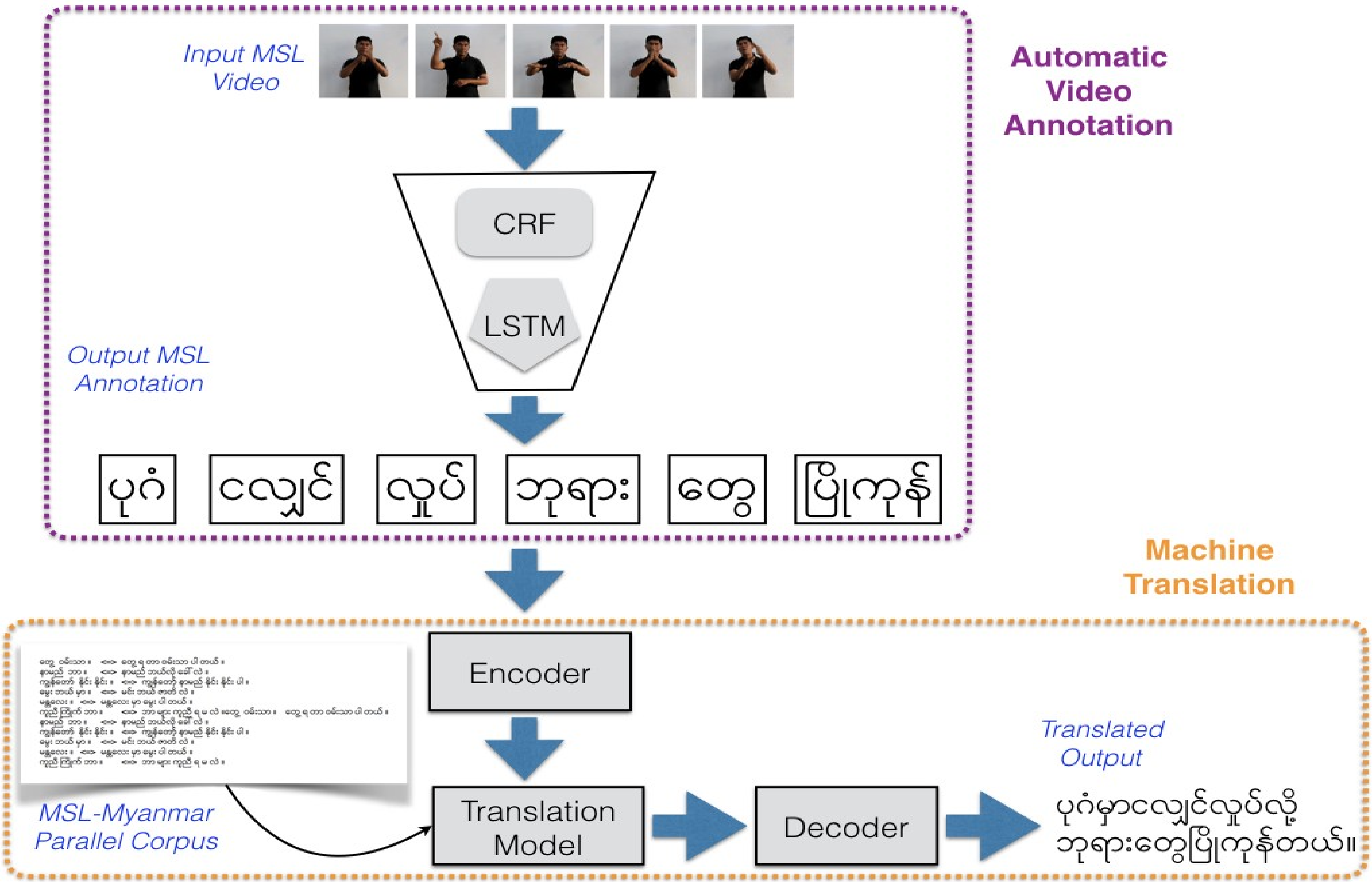



Figure 1. Overview of proposed machine translation system for Myanmar sign language

# Proposed System Design (con't)

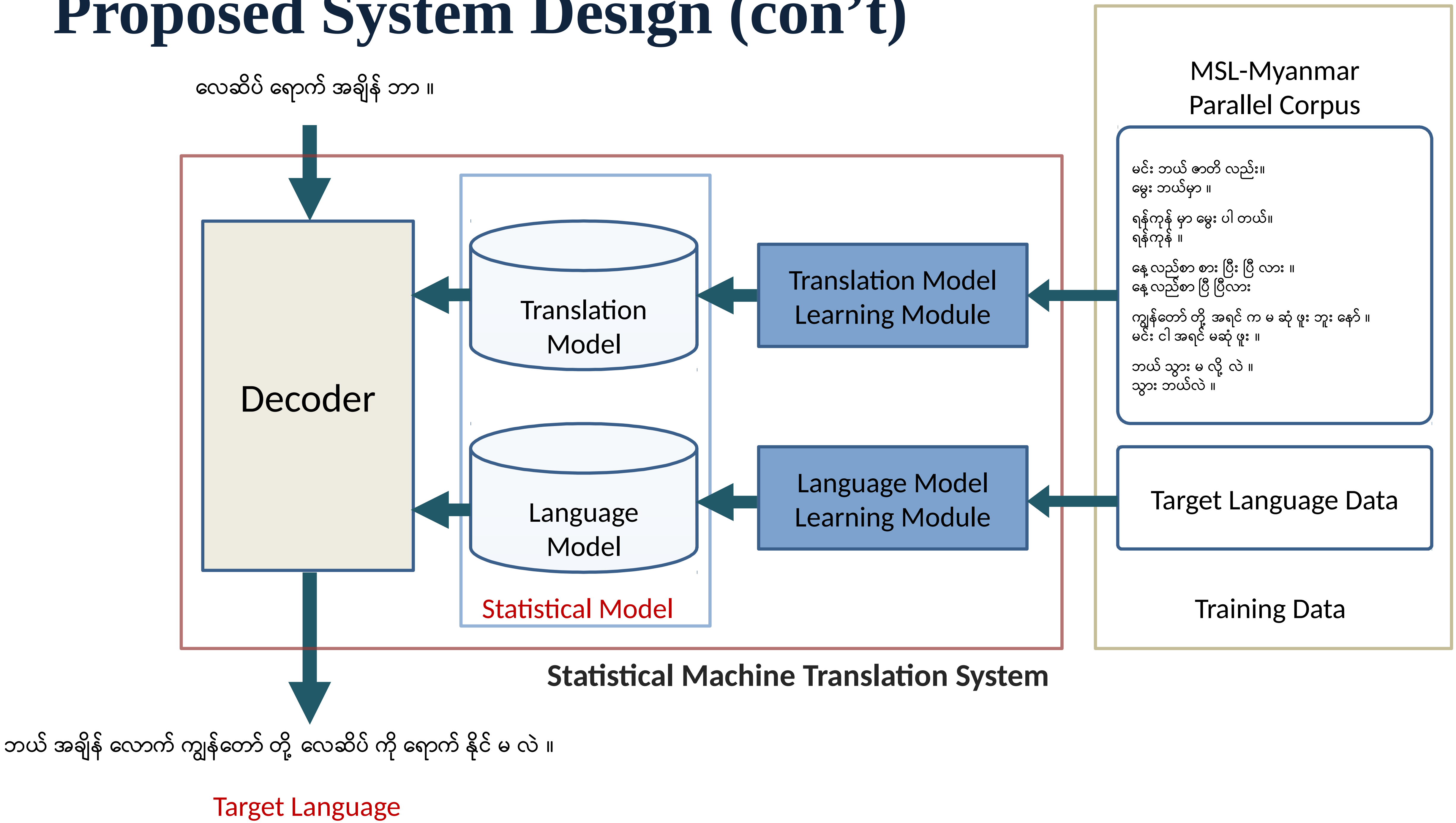



Figure 2. Statistical MT system of the proposed system

# Proposed System Design (con’t)

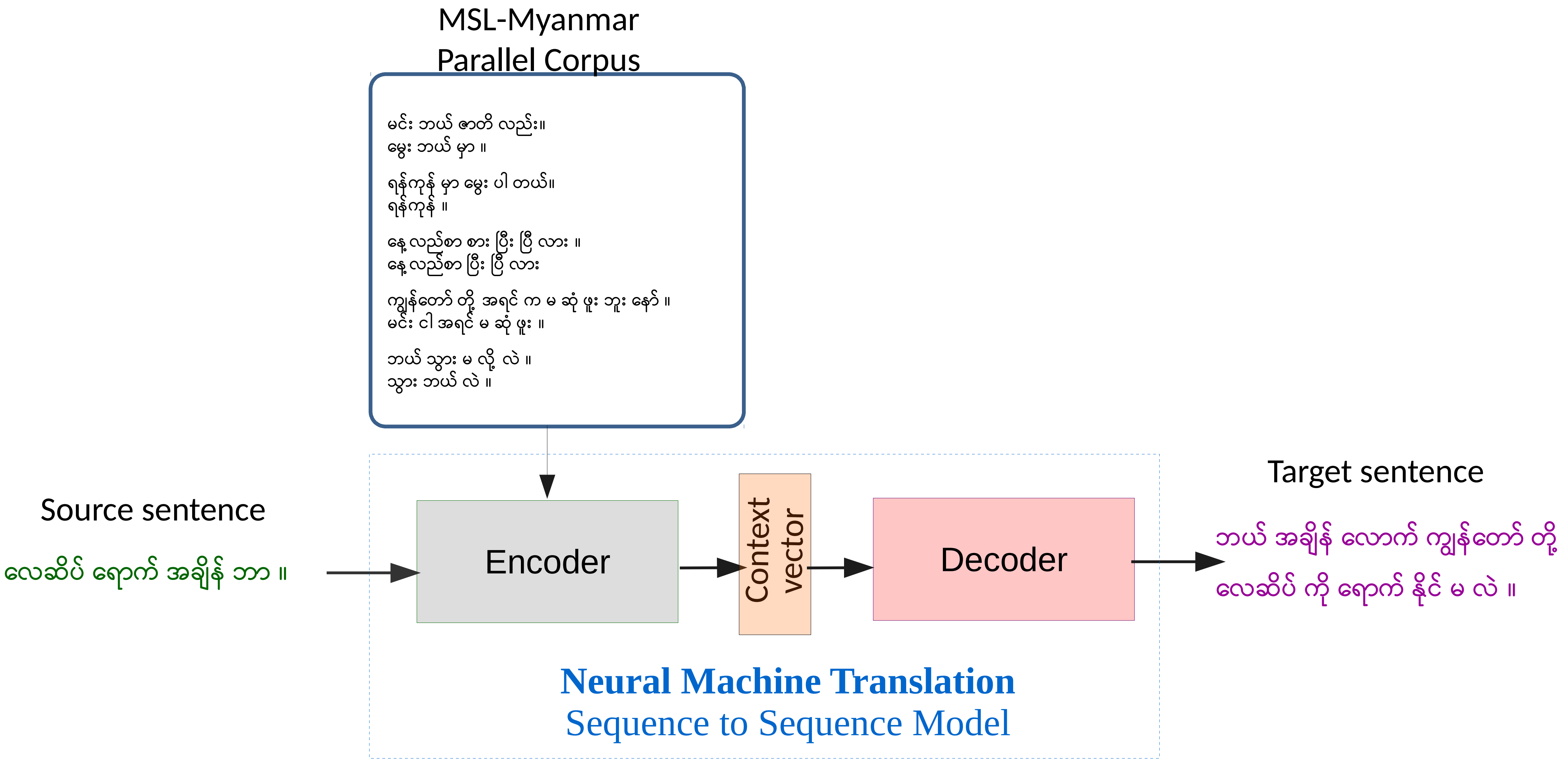



Figure 3. Neural MT system of the proposed system

# Proposed System Design (con't)

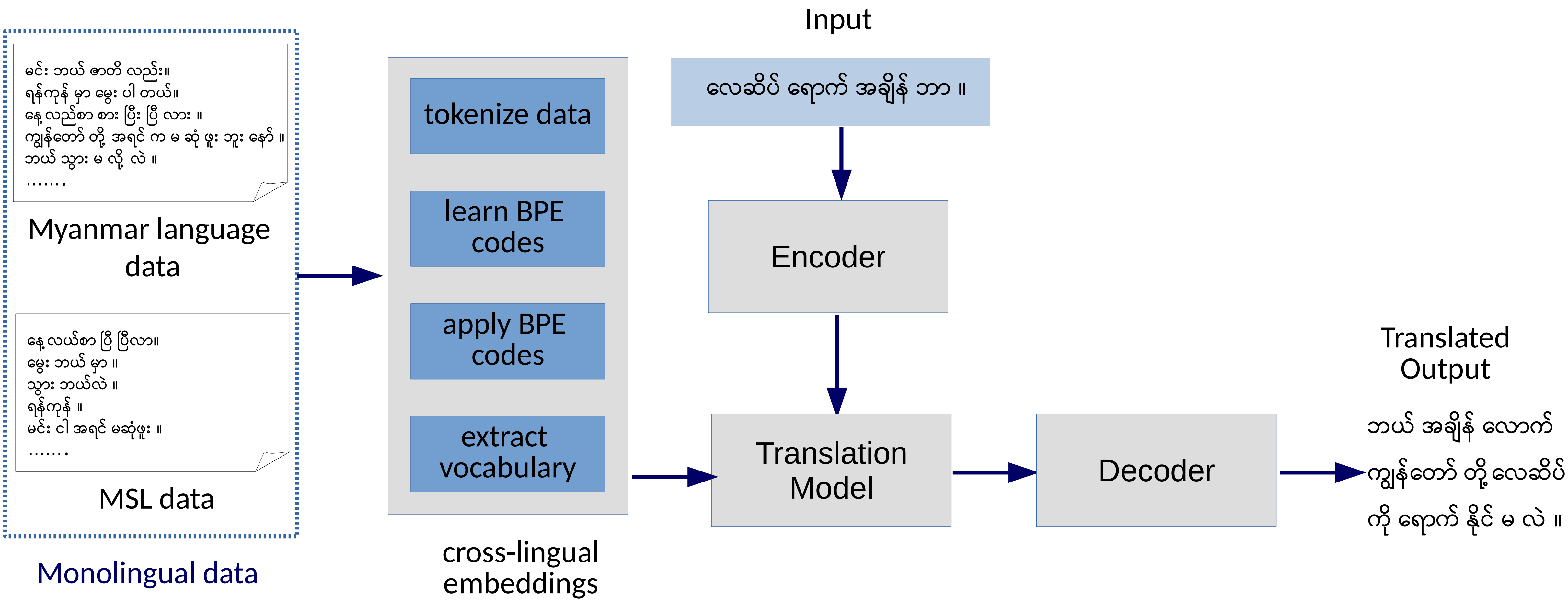



Figure 4. Unsupervised Neural MT system of the proposed system

# MSL and Myanmar Language

Myanmar : မိသားစု မှာ လူ ဘယ် နှစ် ယောက် ရှိ သလဲ ။

MSL : မိသားစု ဘယ်လောက်

English : How many people are there in your family?

**Manual Features**
- handshape,
- movement,
- Location,
- palm orientation

**Non-Manual Features**

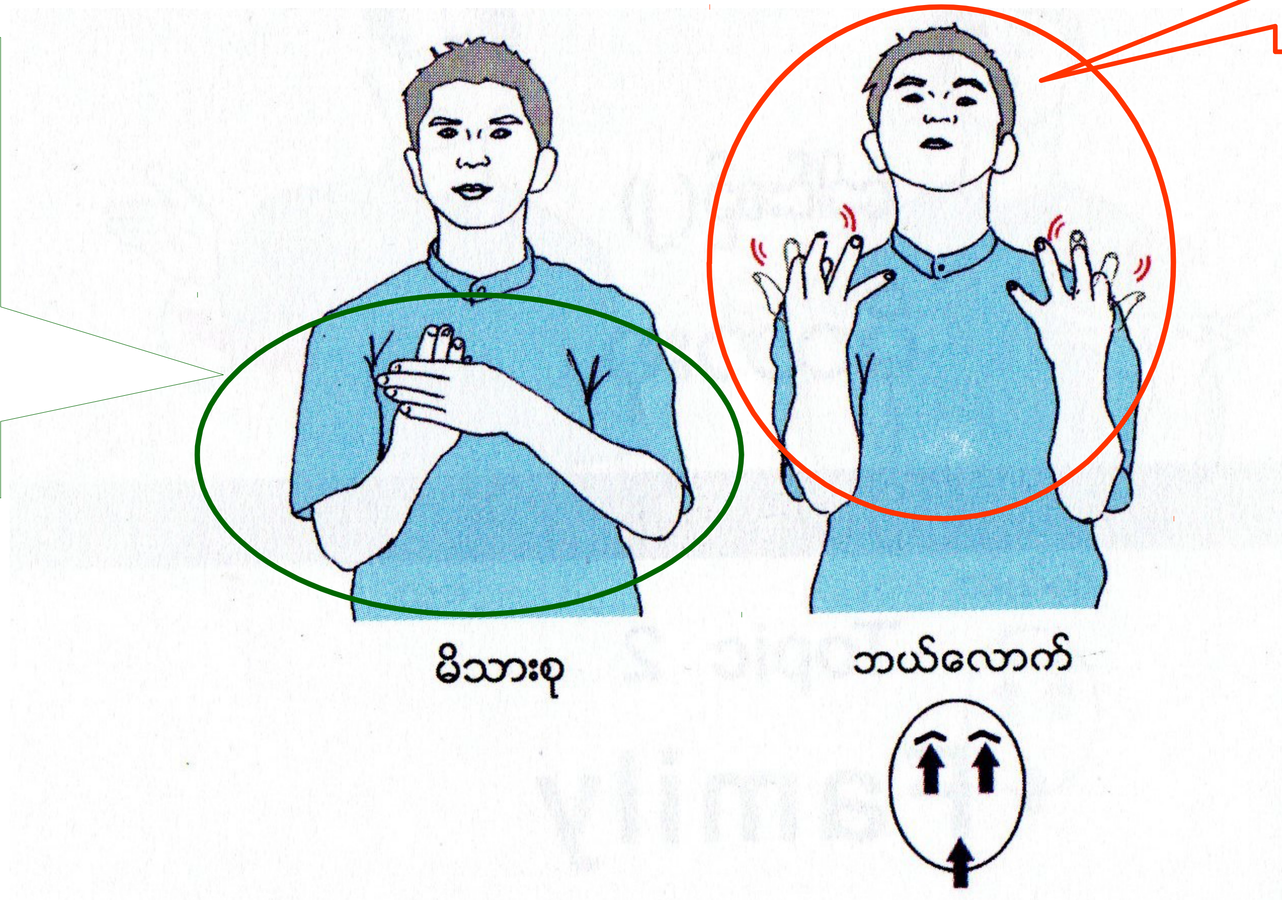



Figure 5. An example of MSL sentence that used non-manual features

# MSL and Myanmar Language (cont.)

•Examples of different grammar, word order and vocabulary between Myanmar language and MSL can be seen in the followings.

| | |
|---|---|
| English | What time do you wake up? |
| Myanmar | ဘယ် အချိန် အိပ်ယာ က ထ သလဲ ။ |
| MSL | အိပ်ယာထ (wake up) အချိန် (time) ဘာလဲ (what) |
| | |
| English | I wake up at six o’clock. |
| Myanmar | မနက် ခြောက် နာရီ မှာ ထ လေ့ ရှိ ပါ တယ် ။ |
| MSL | မနက် (morning) နာရီ (o’clock) ခြောက် (six) |
| | |
| English | Daughter-in-law |
| Myanmar | ချွေးမ |
| MSL | သား (son) လက်ထပ် (marries) မိန်းကလေး (girl) |

# Corpus Preparation

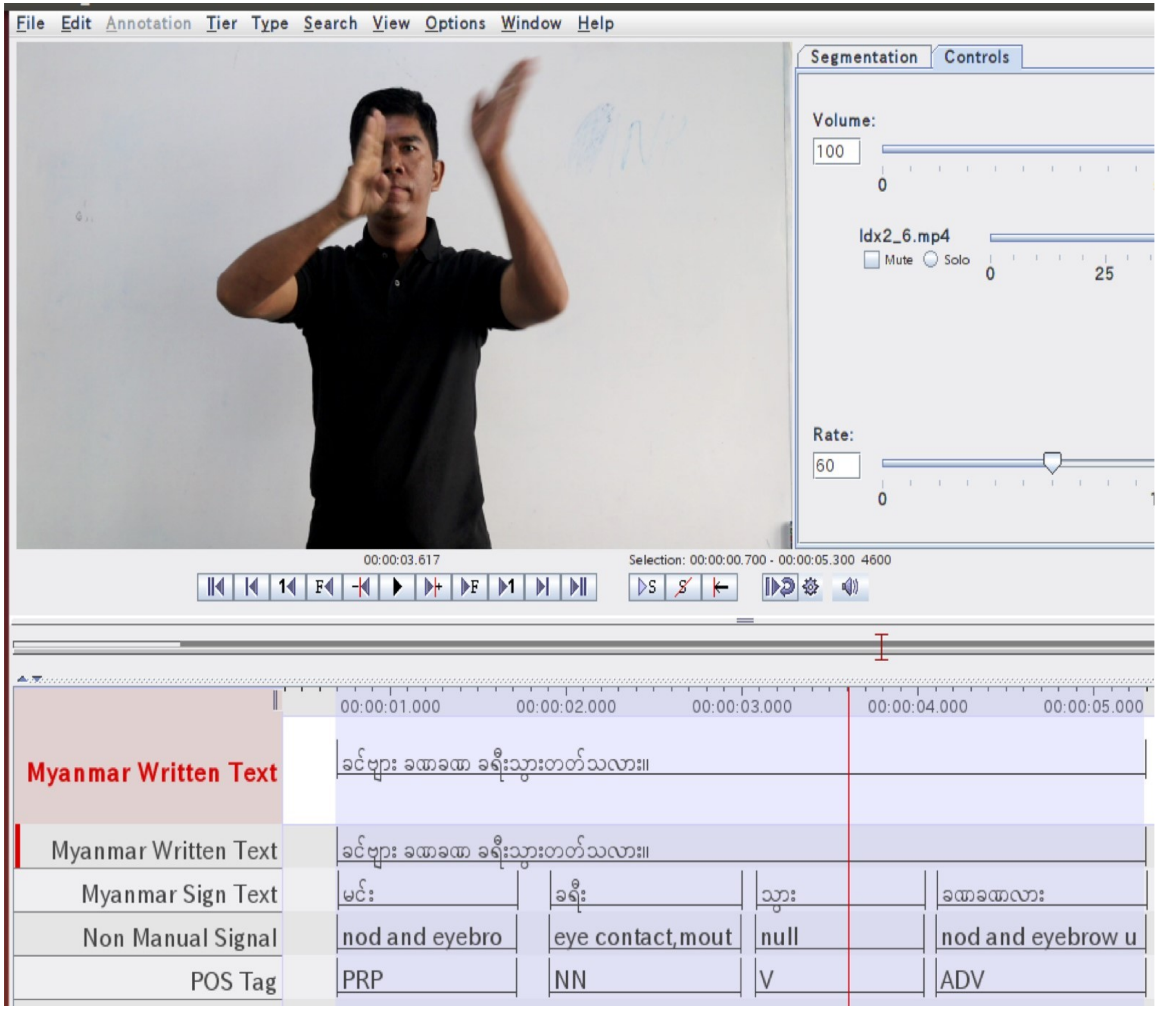



Figure 6. An example of video annotation with ELAN (Eudico Linguistic Annotator)

# Corpus Preparation (cont.)
## (Segmentation)

- **Word segmentation for Myanmar sentence –** applied the word segmentation rules proposed by Win Pa Pa, et al (2015)

  eg. သူ ဝဏ္ဏ နဲ့ စေ့စပ်ကြောင်းလမ်း ထား တယ် ။

  – **Sign Unit based word segmentation for MSL –** based on the meaningful sign unit words

  eg. သူ စေ့စပ် ကောင်လေး နာမည် စာလုံးပေါင်း ဝ ဏ ဏ ။

- **Syllable Segmentation -** used Regular Expression (RE) based Myanmar syllable segmentation tool named "sylbreak"

  eg. သူ ဝဏ္ဏ နဲ့ စေ့ စပ် ကြောင်း လမ်း ထား တယ် ။

- **SentencePiece -** (Taku Kudo and John Richardson, 2018)

  eg. ▁သူ ဝ ဏ္ ဏ နဲ့ စေ့စပ် ကြောင်း လမ်း ထားတယ် ။

- **Byte-Pair-Encoding (BPE) -** (Sennrich et al., 2016)

  eg. သူ ဝ@@ ဏ္@@ ဏ နဲ့ စေ့စ@@ ပ်@@ ကြ@@ ောင်း@@ လမ်း ထား တယ် ။

# Experimental Methodology

- Statistical Machine Translation
  - PBSMT, HPBSMT, OSM
- Neural Machine Translation
  - RNN, Transformer, CNN
- Unsupervised Machine Translation
  - Unsupervised NMT

# Experimental Methodology (cont.)
## (Statistical Machine Translation)

Let ***f*** be any text in source language

$$\hat{e} = argmax_e \, \mathrm{P}(e \mid f)$$

Applying the Bayes' rule,

$$\mathrm{P}(e \mid f) = \frac{\mathrm{P}(e)\,\mathrm{P}(f \mid e)}{\mathrm{P}(f)}$$

$$argmax_e \, \mathrm{P}(e \mid f) = argmax_e \, \mathrm{P}(f \mid e)\,\mathrm{P}(e)$$

# Experimental Methodology (cont.)
## (Statistical Machine Translation)

$$P(e|f) = argmax_e P(f|e)P(e)$$

**Language Model** $P(e)$

- Takes care of fluency in the target language
- Data: corpora in the target language

**Translation Model** $P(f|e)$

- Lexical correspondence between languages
- Data: aligned corpora in source and target languages

*argmax*

- Search done by the decoder

# Experimental Methodology (cont.)
## (Statistical Machine Translation)

**Language Model**

$$P(e|f) = argmax_e P(f|e) P(e)$$

Estimation of how probable a sentence is.

Naïve estimation on a corpus with N sentences:

Frequentist probability of a sentence e:

$$P(e) = \frac{N_e}{N_{sentences}}$$

**Problem:**

Long chains are difficult to observe in corpora.

⇒ Long sentences may have zero probability!

# Experimental Methodology (cont.)
## (Statistical Machine Translation)

**The n-gram approach**

The language model assigns a probability ***P(e)*** to a sequence of words ***e*** $\Rightarrow \{w_1, \ldots, w_m\}$.

$$P(w_1, \ldots, w_m) = \prod^{m} P(w_i | w_{i-(n-1)}, \ldots, w_{i-1})$$

- The probability of a sentence is the product of the conditional probabilities of each word $w_i$ given the previous ones.
- Independence assumption: the probability of $w_i$ is only conditioned by the *n* previous words.

# Experimental Methodology (cont.)
## (Statistical Machine Translation)

Example, a 4-gram model

$e$ : တစ် နာရီ ကို ၆ ဒေါ်လာ

$$P(e) = P(\text{တစ်}) \times P(\text{နာရီ} \mid \text{တစ်}) \times P(\text{ကို} \mid \text{တစ် နာရီ}) \times P(\text{၆} \mid \text{တစ် နာရီ ကို}) \times P(\text{ဒေါ်လာ} \mid \text{နာရီ ကို ၆})$$

where, for each factor,

$$P(\text{၆} \mid \text{တစ် နာရီ ကို}) = \frac{N_{(\text{တစ် နာရီ ကို ၆})}}{N_{(\text{တစ် နာရီ ကို})}}$$

# Experimental Methodology (cont.)
## (Statistical Machine Translation)

**Translation Model**

$$P(e|f)=argmax_e P(f|e)P(e)$$

Estimation of the lexical correspondence between languages.

How can be P(f |e) characterized?

MSLsentence ⇒ NULL တစ်ပတ် ၇ရက် အလုပ် နာရီပေါင်း ဘယ်လောက် ။

Myanmar sentence ⇒ တစ် ပတ် မှာ ဘယ် နှစ် နာရီ အလုပ် လုပ် သလဲ ။

One should at least model for each word in the source language:

- Its translation,
- the number of necessary words in the target language,
- the position of the translation within the sentence,
- and, besides, the number of words that need to be generated from scratch.

# Experimental Methodology (cont.)
## (Statistical Machine Translation)

**Word-based models: the IBM models**

They characterise $P(f \mid e)$ with 4 parameters: $\boldsymbol{t}$, $\boldsymbol{n}$, $\boldsymbol{d}$ and $\boldsymbol{p_1}$ .

⇒ **Lexical probability *t***
t(နာရီပေါင်း | နာရီ) : the prob. that နာရီပေါင်း translates into နာရီ.

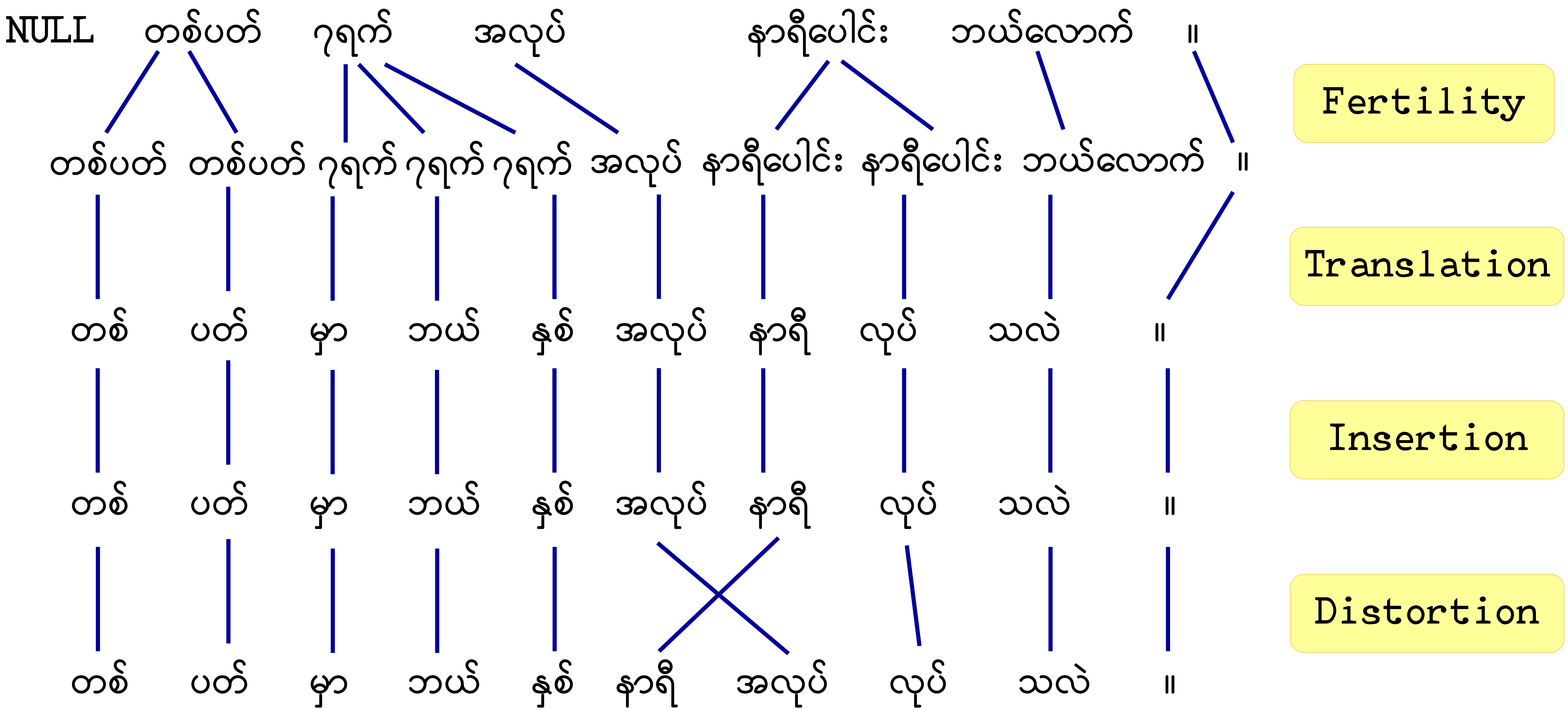
⇒ **Fertility *n***
n(3|ဂျက်): the prob. that ဂျက် generates 3 words.

⇒ **Distortion *d***
$d(j \mid i, m, n)$: the prob. that the word in the $j$ position generates a word in the $i$ position. $m$ and $n$ are the length of the source and target sentences.

⇒ **Probability $\boldsymbol{p_1}$**
the prob. that the spurious word is generated (from NULL).

# Experimental Methodology (cont.)
## (Statistical Machine Translation)

The translation model P(f | e)



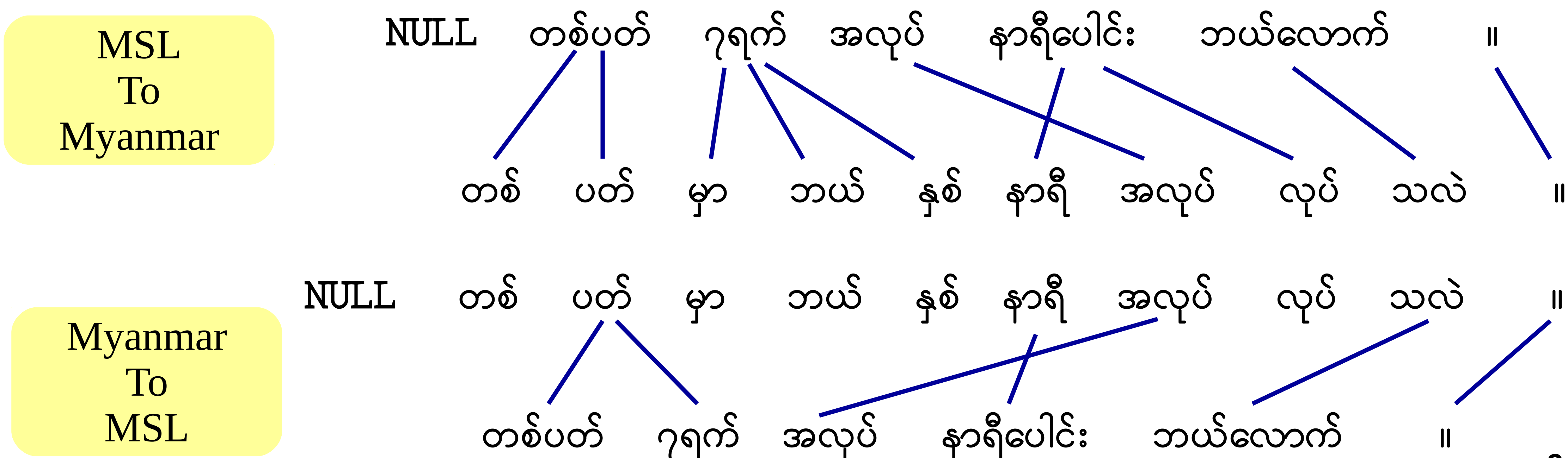
# Experimental Methodology (cont.)
## (Statistical Machine Translation)

The translation model $P(f \mid e)$

**Alignment's asymmetry**

The definitions in IBM models make the alignments asymmetric

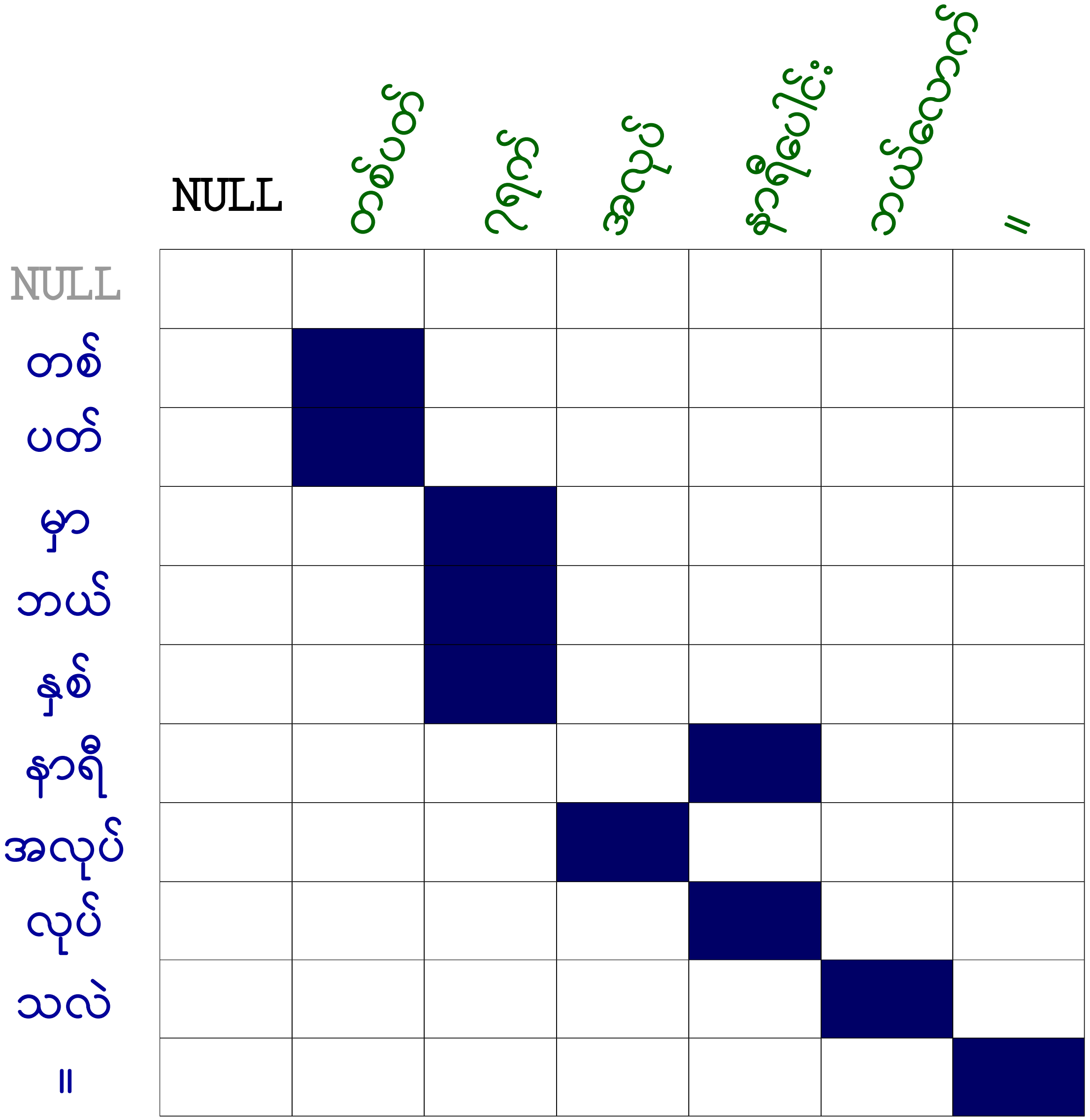
- each target word corresponds to only one source word, but the opposite is not true due to the definition of fertility.



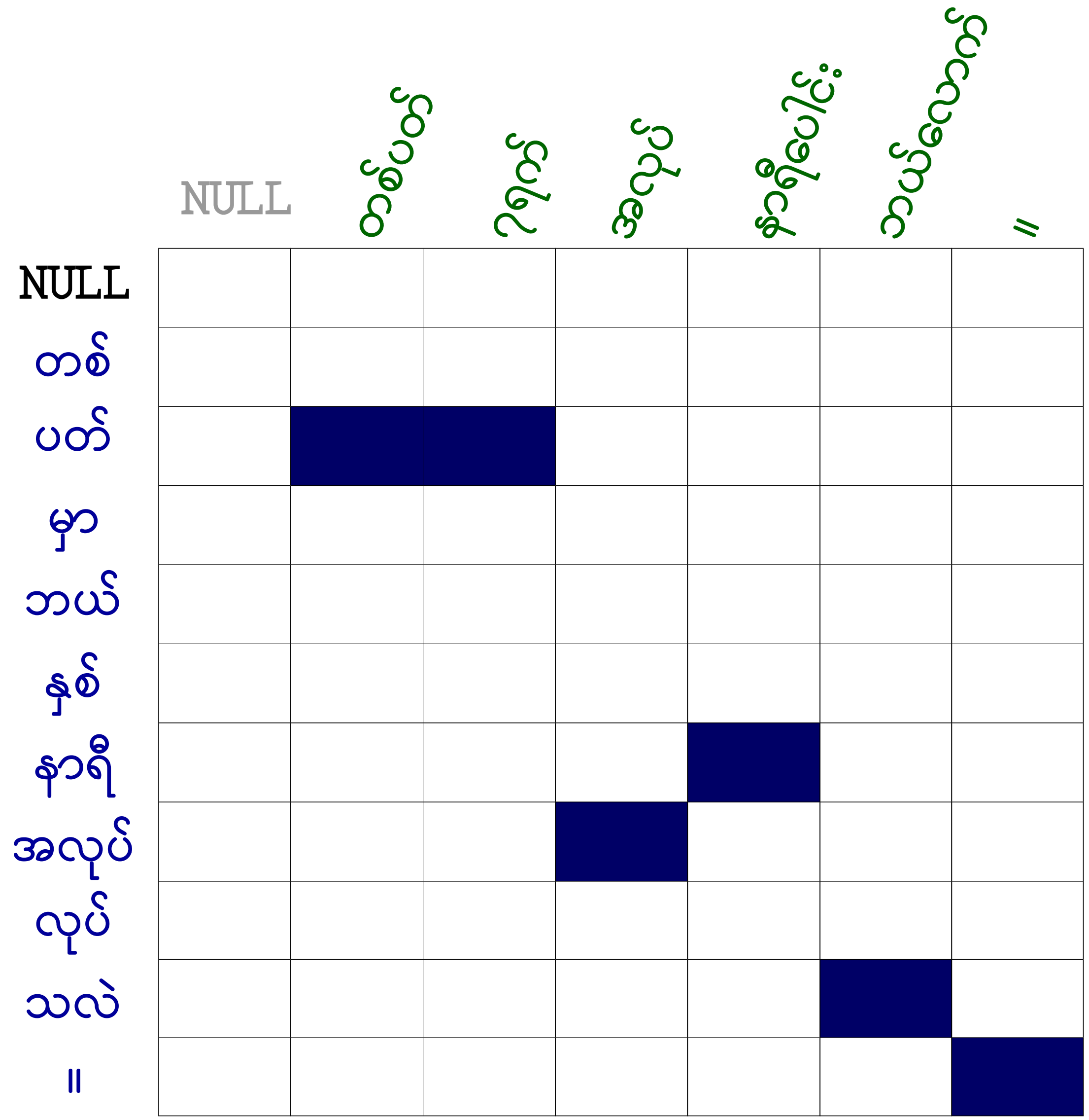
# Experimental Methodology (cont.)

## (Statistical Machine Translation)

The translation model $P(f \mid e)$




MSL to Myanmar




Myanmar to MSL

# Experimental Methodology (cont.)
## (Phrase Based SMT)

- PBSMT translation model is based on phrasal units.
- It typically gives better translation performance than word-based models.
- It consists of
  - phrase-pair probabilities extracted from corpus
  - reordering model
  - algorithm to extract the phrases to build a phrase-table

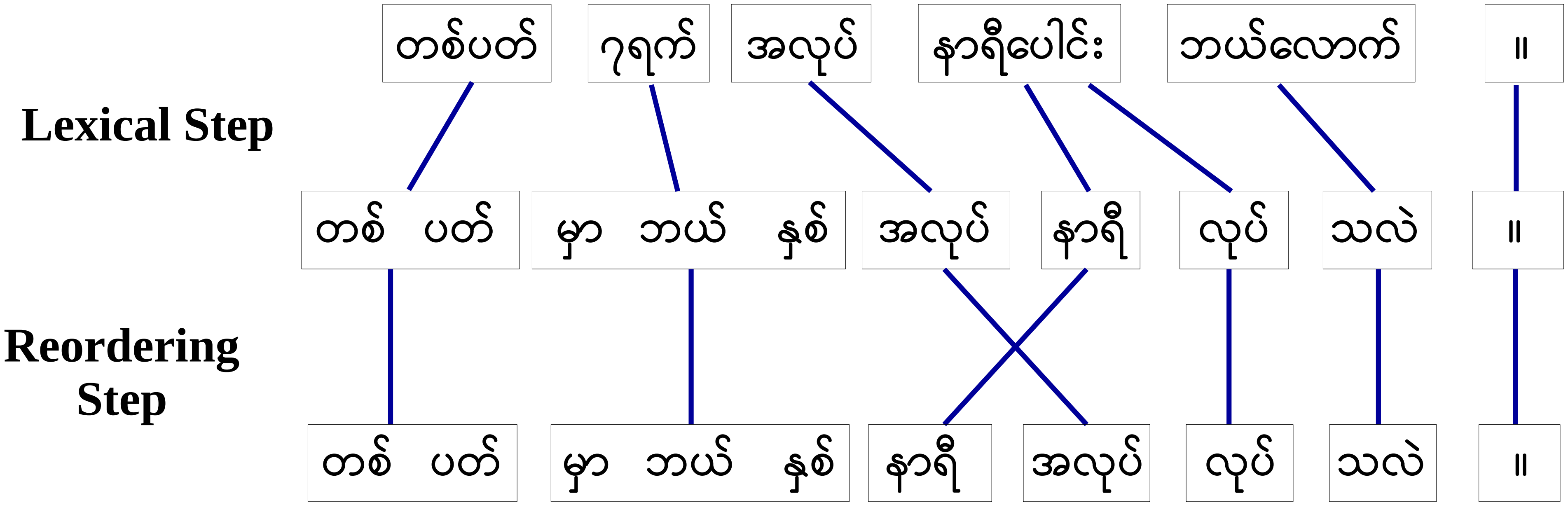
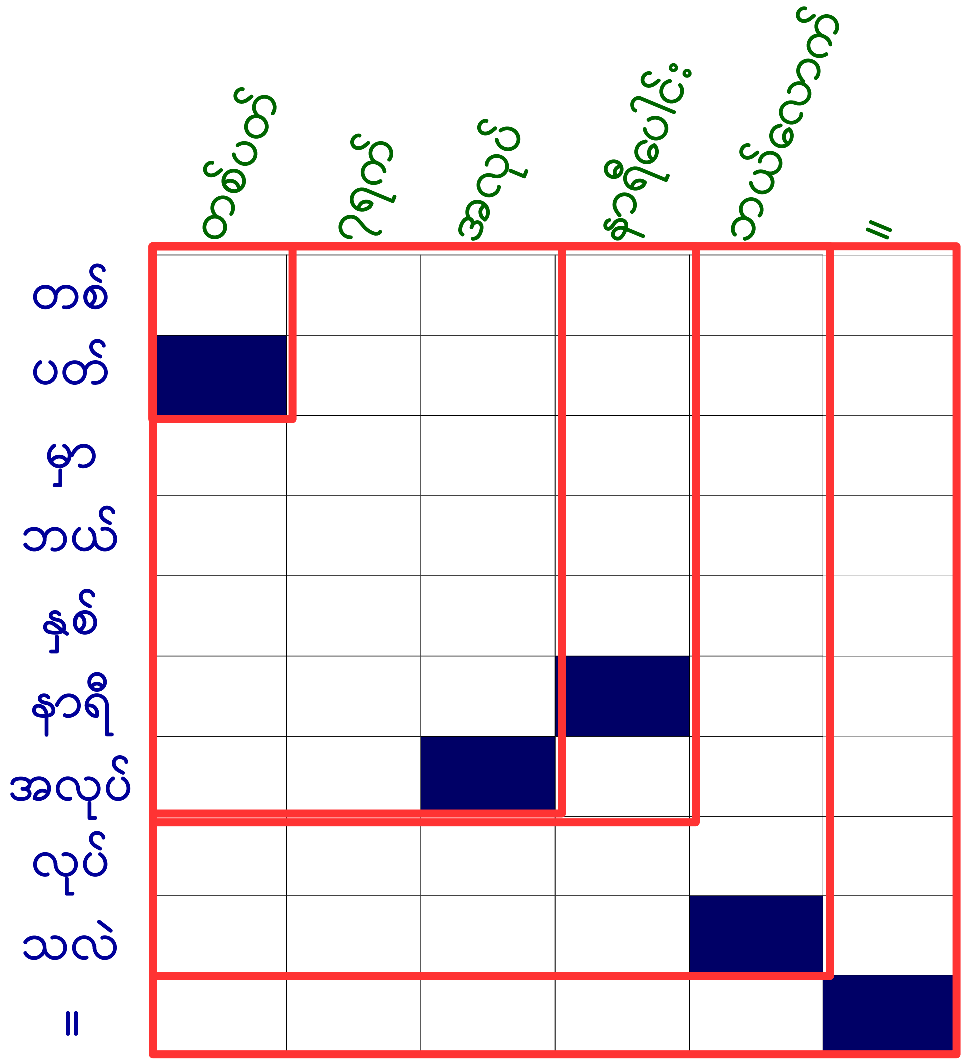



Figure 7. Phrase-based translation

# Experimental Methodology (cont.)
## (Phrase Based SMT)

**Phrase extraction** through an example:

Intersection



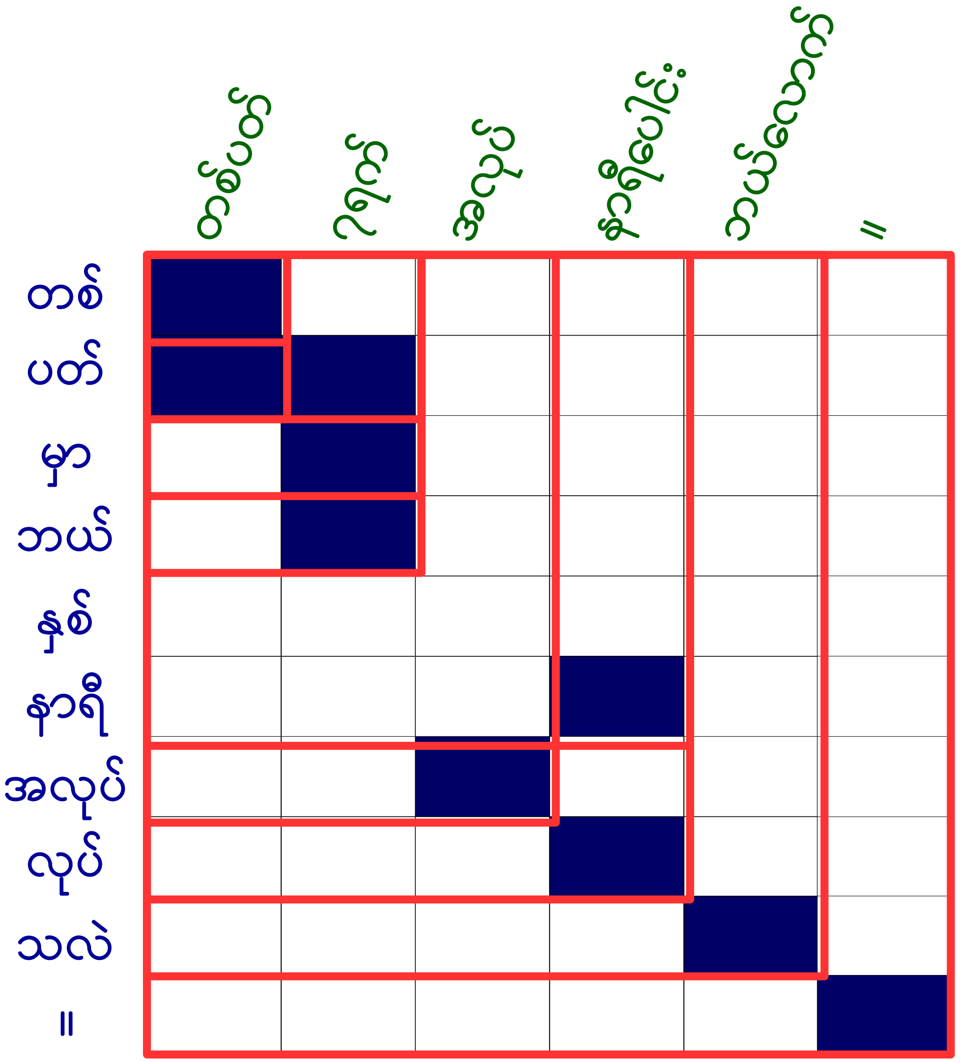

(တစ်ပတ်, တစ် ပတ်)
(တစ်ပတ် ၇ရက် အလုပ်, တစ် ပတ် မှာ ဘယ် နှစ် နာရီ အလုပ်)
(၇ရက် အလုပ် နာရီပေါင်း, တစ် ပတ် မှာ ဘယ် နှစ် နာရီ အလုပ်)
(၇ရက် အလုပ် နာရီပေါင်း ဘယ်လောက်, တစ် ပတ် မှာ ဘယ် နှစ် နာရီ အလုပ် လုပ် သလဲ )
(၇ရက် အလုပ် နာရီပေါင်း ဘယ်လောက် ။, တစ် ပတ် မှာ ဘယ် နှစ် နာရီ အလုပ် လုပ် သလဲ ။)
(အလုပ်, အလုပ်) (အလုပ် နာရီပေါင်း, နာရီ အလုပ်)
(အလုပ် နာရီပေါင်း ဘယ်လောက်, နာရီ အလုပ် လုပ် သလဲ)
(အလုပ် နာရီပေါင်း ဘယ်လောက် ။, နာရီ အလုပ် လုပ် သလဲ ။)
(နာရီပေါင်း , နာရီ) (နာရီပေါင်း ဘယ်လောက် , နာရီ အလုပ် လုပ် သလဲ)
(ဘယ်လောက်, သလဲ) (ဘယ်လောက် ။, သလဲ ။) (။, ။)

14 phrases

# Experimental Methodology (cont.)
## (Phrase Based SMT)

**Phrase extraction** through an example:

**Union**




(တစ်ပတ်, တစ်) (တစ်ပတ်, တစ် ပတ်)

(တစ်ပတ် ၇ရက်, တစ် ပတ်)(တစ်ပတ် ၇ရက်, တစ် ပတ် မှာ)

(တစ်ပတ် ၇ရက်, တစ် ပတ် မှာ ဘယ်)

(တစ်ပတ် ၇ရက် အလုပ်, တစ် ပတ် မှာ ဘယ် နှစ် နာရီ အလုပ်)

(တစ်ပတ် ၇ရက် အလုပ် နာရီပေါင်း, တစ် ပတ် မှာ ဘယ် နှစ် နာရီ)

(တစ်ပတ် ၇ရက် အလုပ် နာရီပေါင်း, တစ် ပတ် မှာ ဘယ် နှစ် နာရီ အလုပ် လုပ်)

(တစ်ပတ် ၇ရက် အလုပ် နာရီပေါင်း ဘယ်လောက်, တစ် ပတ် မှာ ဘယ် နှစ် နာရီ အလုပ် လုပ် သလဲ)

(တစ်ပတ် ၇ရက် အလုပ် နာရီပေါင်း ဘယ်လောက်, တစ် ပတ် မှာ ဘယ် နှစ် နာရီ အလုပ် လုပ် သလဲ)

(၇ရက်, တစ် ပတ်)(၇ရက်, တစ် ပတ် မှာ) . . .

. . . (ဘယ်လောက်, သလဲ)(ဘယ်လောက် ။, သလဲ ။)(။, ။)

31 phrases

# Experimental Methodology (cont.)
## (Phrase Based SMT)

**Drawbacks:**

**Handling of Non-local Dependencies**. Dependencies across phrase boundaries are ignored because of the strong phrasal independence assumption.

**Weak Reordering Model.** Cannot properly handle long-range jumps.

**Hard Distortion Limit.** The lexicalized reordering model fails to filter out bad large-scale reorderings effectively (Koehn 2010).

**Spurious Phrasal Segmentation.** A problem with the phrase-based model is that there is no unique correct phrasal segmentation of a sentence.

# Experimental Methodology (cont.)
## (Hierarchical Phrase Based SMT)

- HPBSMT approach is a model based on synchronous context-free grammar.
- The advantage of this technique offers over the phrase-based approach is that the hierarchical structure is able to represent the word re-ordering process.

ခင်ဗျား ထပ် [X] ||| မင်း ကြီး [X]
ခင်ဗျား ထပ် [X] ||| အသက် မင်း ကြီး [X]
ခင်ဗျား ထပ် [X] [X] [X] ||| မင်း ကြီး [X] [X] [X]
ခင်ဗျား ထပ် [X] [X] [X] ||| အသက် မင်း ကြီး [X] [X] [X]
ခင်ဗျား ထပ် အသက် ငယ် တယ် [X] ||| မင်း ကြီး ငါ ငယ် [X]
ခင်ဗျား ထပ် အသက် ငယ် တယ် [X] ||| အသက် မင်း ကြီး ငါ ငယ် [X]

Figure 8. Some examples of hierarchical phrase-based grammar between MWT and MSL phrases

# Experimental Methodology (cont.)
## (Operation Sequence Model)

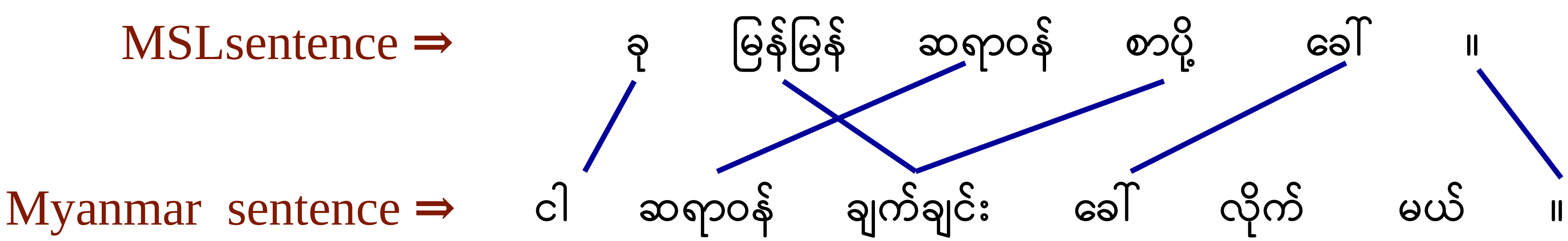
- OSM model which combines the benefits of two state-of-the-art SMT frameworks named n-gram based SMT and phrase-based SMT.
- OSM model uses **five translation** and **three reordering operations**, which are repeatedly applied in a sequence.

**Translation Operations**

- Generate (X,Y)
- Continue Source Cept
- Generate Source Only (X)
- Generate Target Only (Y)
- Generate Identical

**Reordering Operations**

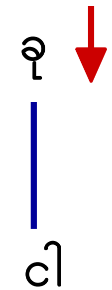
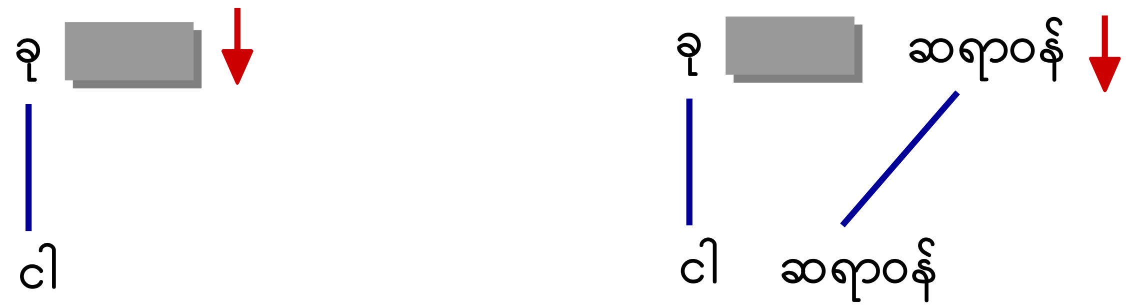
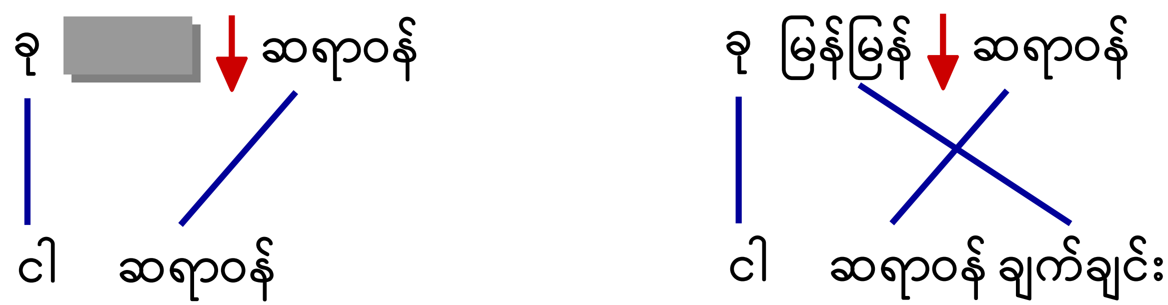
- Insert Gap
- Jump Back (W)
- Jump Forward




Discontinuous cept translation

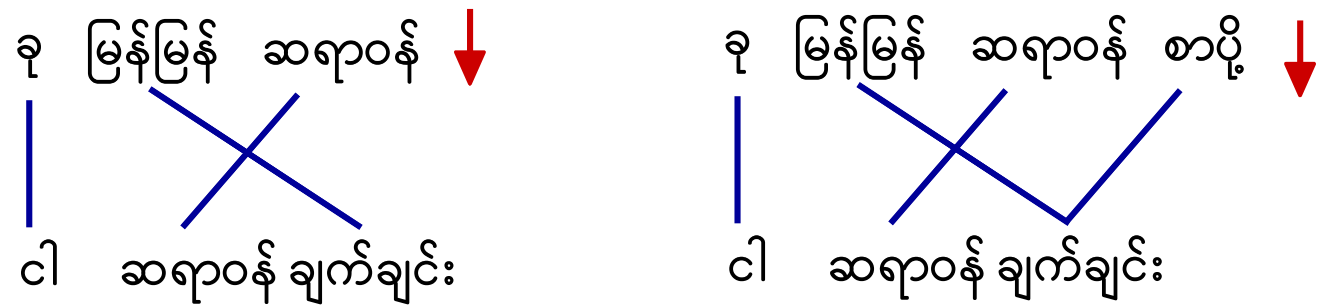
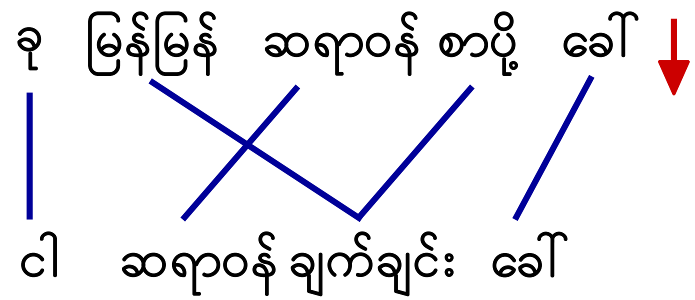
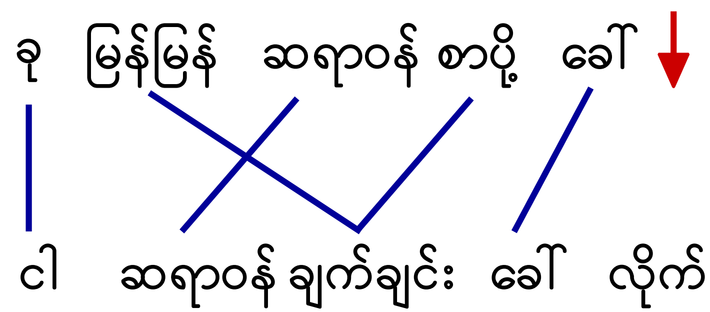
# Experimental Methodology (cont.)
## (Operation Sequence Model)

**Discontinuous cept translation example**

| Operations Sequence | Generation |
| --- | --- |
| Generate(ခု, ငါ) |  |
| Insert Gap<br>Generate Identical |  |
| Jump Back (1)<br>Generate(မြန်မြန် စာပို့, ချက်ချင်း) |  |

# Experimental Methodology (cont.)
## (Operation Sequence Model)

**Discontinuous cept translation example**

| Operations Sequence | Generation |
|---|---|
| Jump Forward<br>Continue Source Cept |  |
| Generate Identical |  |
| Generate Target Only (လိုက်) |  |

# Experimental Methodology (cont.)
## (Operation Sequence Model)

**Discontinuous cept translation example**

| Operations Sequence | Generation |
|---|---|
| Generate Target Only (မယ်) | ခု မြန်မြန် ဆရာဝန် စာပို့ ခေါ်<br>ငါ ဆရာဝန် ချက်ချင်း ခေါ် လိုက် မယ် |
| Generate Identical | ခု မြန်မြန် ဆရာဝန် စာပို့ ခေါ် ။<br>ငါ ဆရာဝန် ချက်ချင်း ခေါ် လိုက် မယ် ။ |

# Experimental Methodology (cont.)
## (Neural Machine Translation)

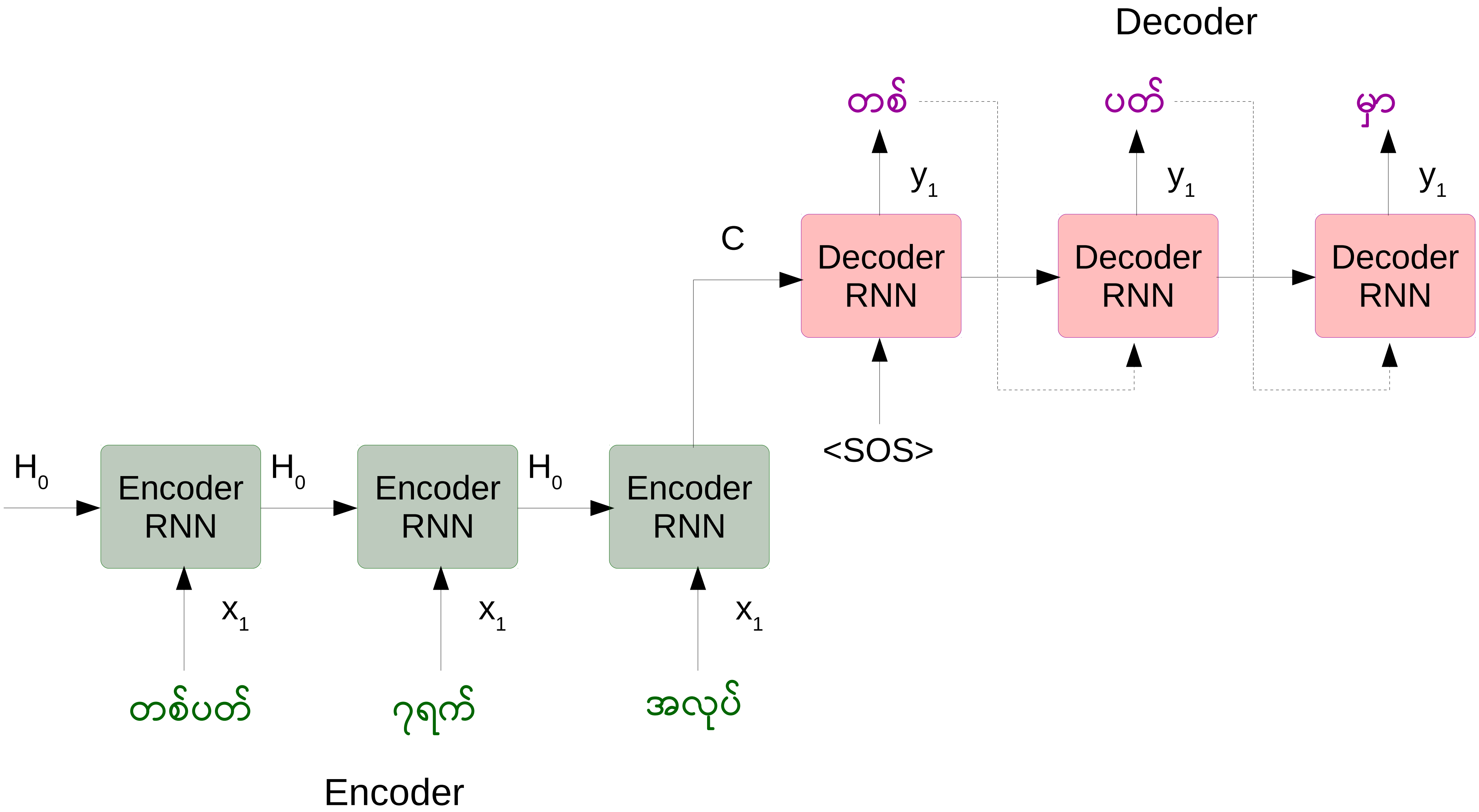



Figure 9. Illustration of Neural Machine Translation

# Experimental Methodology (cont.)
## (Neural Machine Translation)

- Both Encoder and Decoder are RNNs
- At every time step in the Encoder, the RNN takes a word vector ($x_i$) from the input sequence and a hidden state ($H_i$) from the previous time step
- The hidden state is updated at each time step
- The hidden state from the last unit is known as the context vector. This contains information about the input sequence
- This context vector is then passed to the decoder and it is then used to generate the target sequence
- If we use the Attention mechanism, then the weighted sum of the hidden states are passed as the context vector to the decoder

# Experimental Methodology (cont.)
## (Neural Machine Translation)

**Advantages**

- End-to-end models (no pipeline of specific tasks)

**Disadvantages**

- Requires bilingual corpus
- Rare word problem

# Experimental Methodology (cont.)
## (NMT - Stacked RNN with Attention)

[Schwenk, 2012, Kalchbrenner and Blunsom, 2013, Sutskever et al., 2014, Bahdanau et al., 2014, Luong et al., 2015]

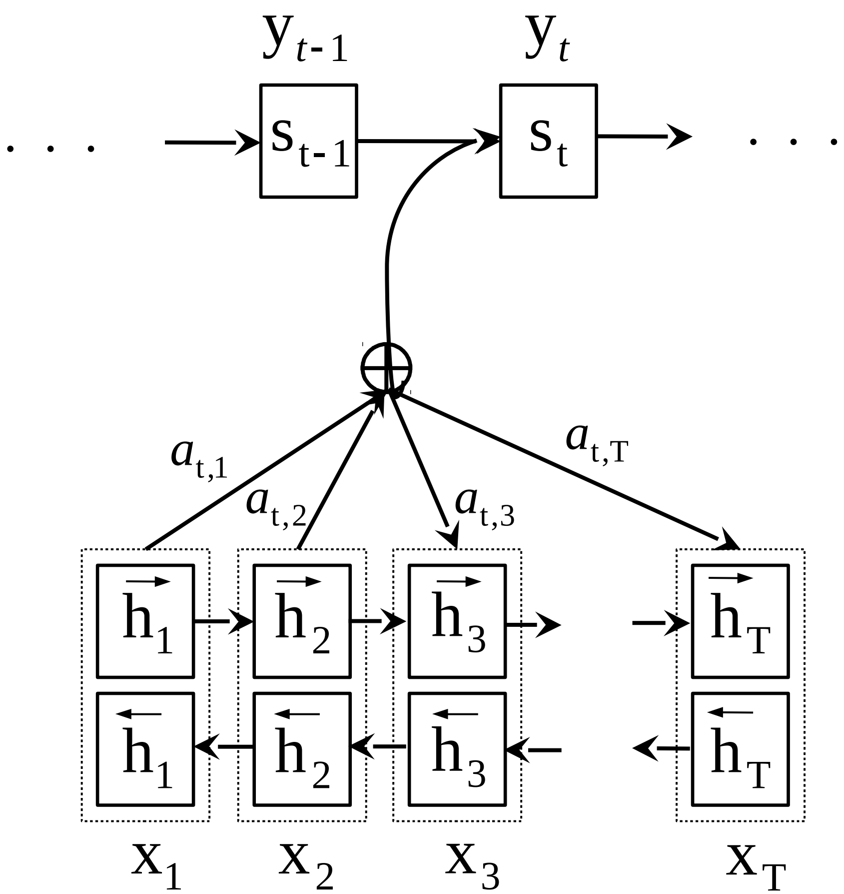



Figure 10. Illustration of a bidirectional RNN as an encoder and decoder that imitates searching through a source sentence decoding a translation.

# Experimental Methodology (cont.)
## (NMT - Stacked RNN with Attention)

- The bidirectional RNN consists of forward and backward RNN’s.
- The forward RNN $\overrightarrow{f}$ reads the input sentence as it is ordered (from $x_1$ to $x_{Tx}$) and calculates a sequence of *forward hidden states* $(\overrightarrow{h}_1, \cdots, \overrightarrow{h}_{T_x})$
- The backward RNN $\overleftarrow{f}$ reads the sequence in the reverse order (from $x_T$ to $x_1$ ), resulting in a sequence of backward hidden states $(\overleftarrow{h}_1, \cdots, \overleftarrow{h}_{T_x})$
- To obtain an annotation for each word $x_j$, concatenate the forward hidden state $\overrightarrow{h}_j$ and the backward one $\overleftarrow{h}_j$ , i.e., $h_j = \left[\overrightarrow{h}_j^\top; \overleftarrow{h}_j^\top\right]^\top$
- In this way, the annotation $h_j$ contains the summaries of both the preceding words and the following words.

# Experimental Methodology (cont.)
## (NMT - Fully Convolutional Models (ConvSeq2Seq))

**[Gehring et al., 2017]**

- Though RNNs have historically outperformed CNNs at language translation tasks, their design has an inherent limitation, which can be understood by looking at how they process information.

- RNNs operate in a strict left-to-right or right-to-left order, one word at a time. This is a less natural fit to the highly parallel GPU hardware that powers modern machine learning.

- The computation cannot be fully parallelized, because each word must wait until the network is done with the previous word.

- In comparison, CNNs can compute all elements simultaneously, taking full advantage of GPU parallelism. They therefore are computationally more efficient.

- Another advantage of CNNs is that information is processed hierarchically, which makes it easier to capture complex relationships in the data.

# Experimental Methodology (cont.)
## (NMT - Self-attentional Transformer)

- The transformer model [Vaswani et al., 2017] uses attention to replace recurrent dependencies, making the representation at time step i independent from the other time steps.
- This allows for parallelization of the computation for all time steps in encoder and decoder.
- The general architecture of a decoder accessing a sequence of source encoder states through an attention mechanism remains the same.
- However, the decoder may use multiple encoder-attention mechanisms in each of its layers.

# Experimental Methodology (cont.)
## (Unsupervised Neural MT)

- Neural Machine Translation has arguably reached human-level performance.
- But, effective training of these systems is strongly dependent on the availability of a large amount of parallel text.
- Because of which supervised techniques have not been so successful in low resource language pairs.
- **Unsupervised Machine Translation** requires only monolingual corpora and is a viable alternative in such cases.
- While it has not been able to outperform supervised learning with lots of parallel resources, it has great potential with low resource language pairs.

# Experimental Methodology (cont.)
## (Unsupervised Neural MT)

(Lample et al., EMNLP 2018 best paper)

**Initialization**

- Initialized shared encoders, with shared cross-lingual BPE embeddings

  i) join the monolingual corpora,

  ii) apply BPE tokenization on the resulting corpus, and

  iii) learn token embeddings (Mikolov et al., 2013) on the same corpus, which are then used to initialize the lookup tables in the encoder and decoder.

# Experimental Methodology (cont.)
## (Unsupervised Neural MT)

**Language Modeling**

- In NMT, language modeling is accomplished via denoising autoencoding.

- **Denoising Auto-Encoding** – The model learns to reconstruct a sentence in a given language from a noisy version of it.
- Critical to add noise to avoid trivial reconstructions

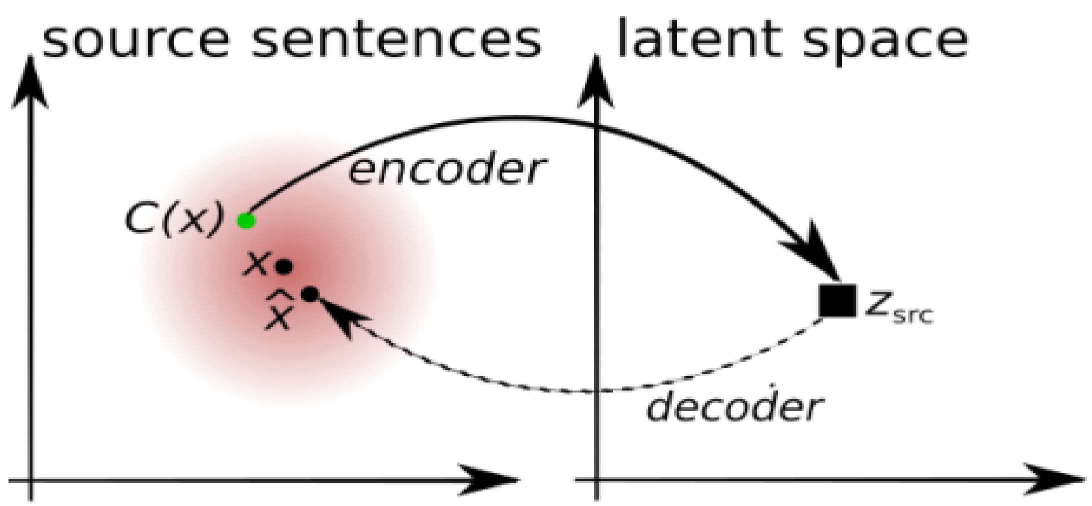



Figure 11. Auto-Encoding – The model is trained to reconstruct a sentence from its noisy version. x is the target, C(x) is the noisy input, x̂ is the reconstruction.

# Experimental Methodology (cont.)
## (Unsupervised Neural MT)

**Iterative Back-translation**

- **Cross-Domain Training** – The model also learns to reconstruct any source sentence given a noisy translation of the same sentence in the target language, and vice versa.

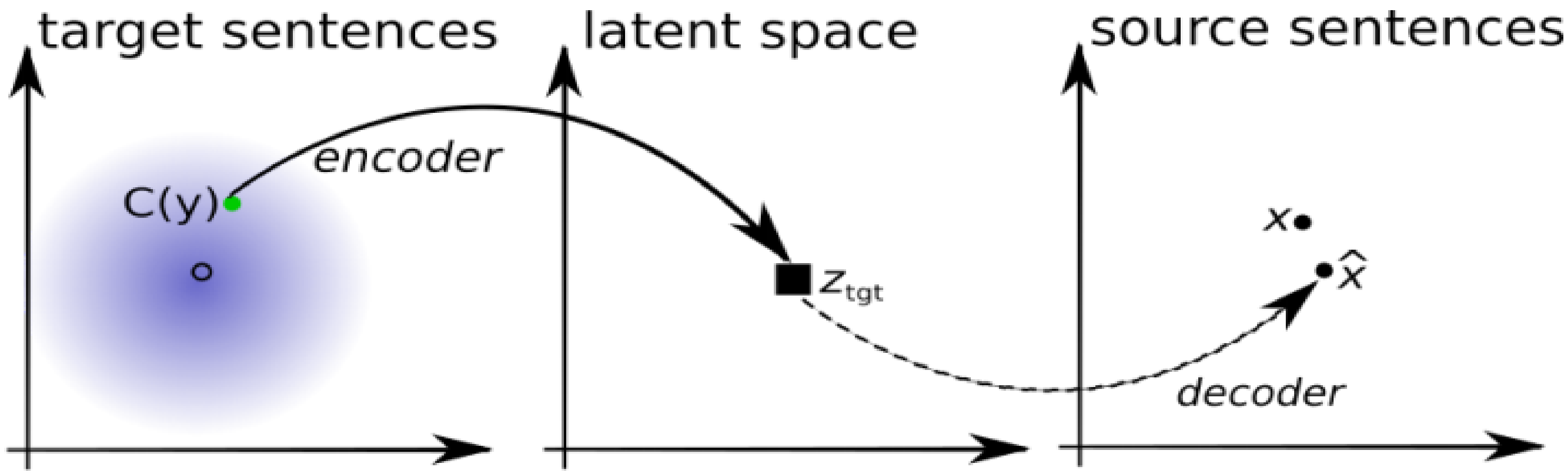



Figure 12. Translation – The model is trained to translate a sentence.
C(y) is the noisy translated sentence generated by the translation model
of the previous iteration, x̂ is the reconstruction of the translated system.

# Experiments and Analysis
## (SMT Experiments)

### Corpus statistics

- 2,510 Myanmar (my) and MSL (sl) parallel sentences of MSL corpus, which 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)
- 6% of sentences are containing Myanmar fingerspelling characters.
- 2,000 sentences were used for training, 310 sentences for development and 200 sentences for evaluation.
- Four types of segmentation pairs (word-word, syllable-syllable, syllable-word, and word-syllable).

### Moses SMT system

- used the PBSMT, HPBSMT and OSM provided by the Moses framework [Kohen et al., 2003]

### Evaluation

- We used the de facto standard automatic evaluation metric Bilingual Evaluation Understudy (**BLEU)** [Och et al., 2000] for the evaluation of the machine translation output.

# Experiments and Analysis (cont.)
## (SMT Experiments)

| Src-Trg | MWT (Syllable) – MSL (Syllable) | | |
|---|---|---|---|
| | PBSMT | HPBSMT | OSM |
| **MWT-MSL** | 34.42 | **35.11** | 34.81 |
| **MSL-MWT** | 33.54 | 33.01 | **34.78** |

Table 1: BLEU scores of syllable-syllable segmentation pair for PBSMT, HPBSMT and OSM

| Src-Trg | MWT (Word) – MSL (Word) | | |
|---|---|---|---|
| | PBSMT | HPBSMT | OSM |
| **MWT-MSL** | 25.80 | **26.42** | 25.38 |
| **MSL-MWT** | 29.77 | 29.70 | **30.38** |

Table 2: BLEU scores of word-word segmentation pair for PBSMT, HPBSMT and OSM.

| Src-Trg | MWT (Syllable) – MSL(Word) | | |
|---|---|---|---|
| | PBSMT | HPBSMT | OSM |
| **MWT-MSL** | 21.02 | **21.96** | 20.55 |
| **MSL-MWT** | 20.93 | 20.18 | **21.21** |

Table 3: BLEU scores of syllable-word segmentation pair for PBSMT, HPBSMT and OSM

| Src-Trg | MWT (Word)- MSL (Syllable) | | |
|---|---|---|---|
| | PBSMT | HPBSMT | OSM |
| **MWT-MSL** | 24.17 | 23.94 | **24.38** |
| **MSL-MWT** | 25.31 | 26.03 | **27.23** |

Table 4: BLEU scores of word-syllable segmentation pair for PBSMT, HPBSMT and OSM

# Experiments and Analysis (cont.)
## (SMT Experiments)

**BLEU scores for MSL Mandalay data only**

| Source - Target | PBSMT | HPBSMT | OSM |
|---|---|---|---|
| MWT→MSL | 13.87 | 13.93 | **13.68** |
| MSL→ MWT | 16.65 | 17.23 | **17.51** |

Training – 4106
Dev – 650
Test - 590

Table 5. BLEU scores of training data (MSL Mandalay data only) for MWT→MSL and MSL→MWT translations tasks

**BLEU scores for Training data (MSL both Mandalay and Yangon)**

| Source - Target | PBSMT | HPBSMT | OSM |
|---|---|---|---|
| MWT→MSL | 14.12 | 14.20 | **14.78** |
| MSL→ MWT | 17.52 | 17.74 | **18.66** |

Training – 4509
Dev – 650
Test - 590

Table 6. BLEU scores of training data (MSL both Mandalay and Yangon) for MWT→MSL and MSL→MWT translations tasks

# Experiments and Analysis (cont.)
## (SMT Error Analysis)

- We analyzed the translated outputs of 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)
- The following example shows WER calculation on the MSL-Myanmar where the word segmentation method. In this example, S=2, D=1, I=3, C=4, and N=7 and WER for whole sentence is equal to 6 / 7 = 86%.

| | | |
|---|---|---|
| **Ref:** ဒီနေ့ အပူချိန် ဒီဂရီ ၃ ၅ ကျော် ။ | | |
| **Hyp:** ဒီ နေ့ နေပူ အပူချိန် ဒီဂရီ ၃၅ ကျော် ကြိုက် ။ | | |
| **WER errors** | | |
| Reference | Hypothesis | Error type |
| | ဒီ | Insertion |
| | နေ့ | Insertion |
| ဒီနေ့ | နေပူ | Substitution |
| ၃ | | Deletion |
| ၅ | ၃၅ | Substitution |
| | ကြိုက် | Insertion |

# Experiments and Analysis (cont.)
## (SMT Error Analysis)

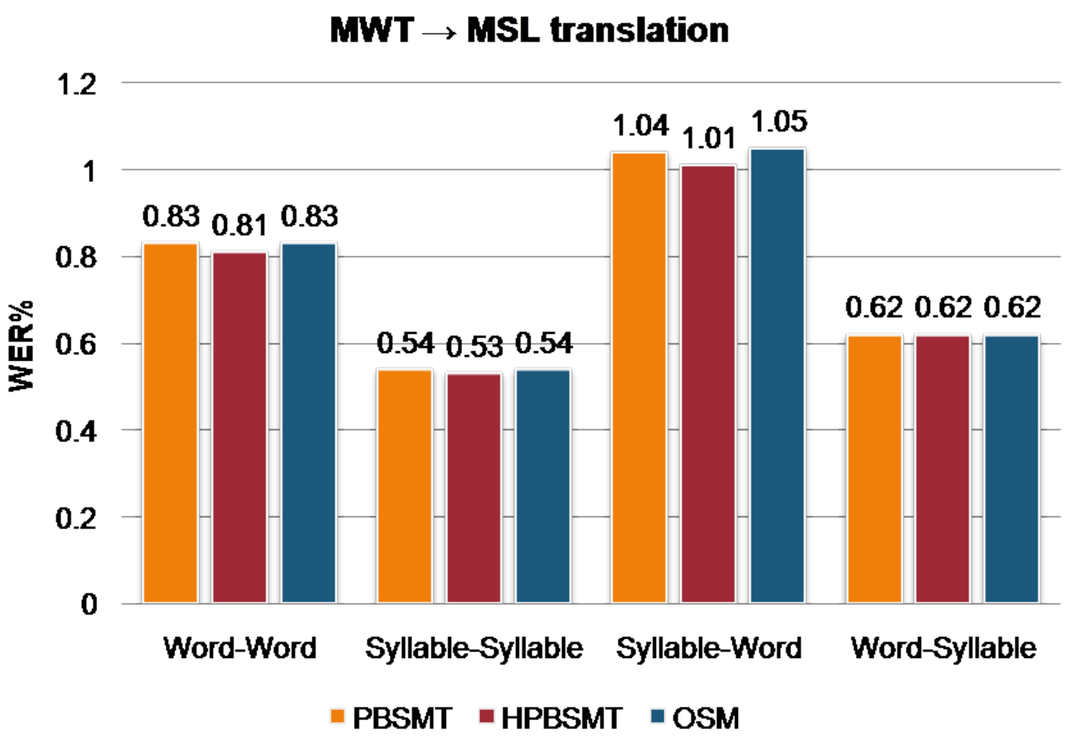



Figure 13: WER of Myanmar written text to Myanmar sign language translation with word-word, syllable-syllable, syllable-word, word-syllable segmentation pairs

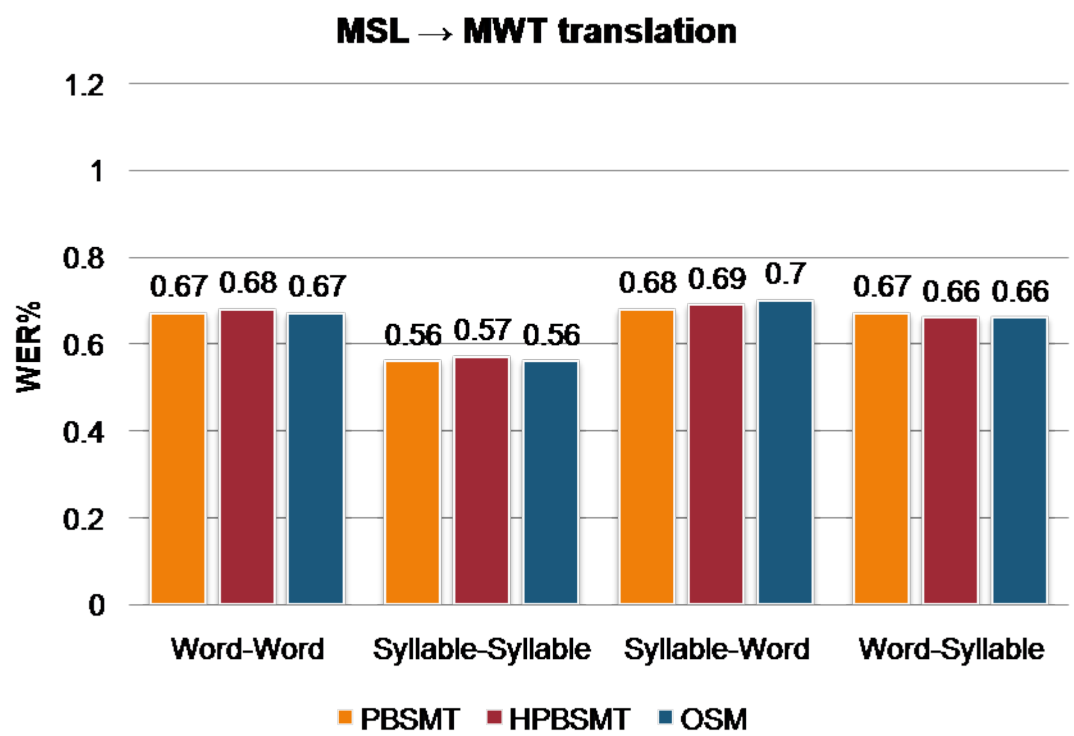



Figure 14: WER of to Myanmar sign language to Myanmar written text translation with word-word, syllable-syllable, syllable-word, word-syllable segmentation pairs

# Experiments and Analysis (cont.)

## (SMT Error Analysis)

- Several reordering errors are found in PBSMT and OSM on the MWT→MSL translation on both syllable and word segmentation.

| **Syllable - Syallable (my-sl)** | |
|---|---|
| **Source:** | သူ မြွေ ကိုက် ခံ ရ လို့ ပါ ။ |
| **Reference:** | သူ မြွေ ကိုက်ခံရ ။ |
| PBSMT hyps: | သူ **ကိုက်** မြွေ ခံ ရ ။ |
| HPBSMT hyp: | သူ မြွေ ကိုက် ခံ ရ ။ |
| OSM hyp: | သူ **ကိုက်**မြွေ ခံ ရ ။ |
| | |
| **Word - Word (my-sl)** | |
| **Source:** | ကျွန်တော် သွား လို့ ရ ပြီ လား ။ |
| **Reference:** | ငါ သွား ရလား ။ |
| PBSMT hyp: | ။ ငါ သွား ပြီးပြီလား ။ |
| HPBSMT hyp: | ငါ သွား ရ လား ။ |
| OSM hyp: | ငါ သွား ။ ပြီးပြီလား ။ |
| | |
| **Syllable - Word (my-sl)** | |
| **Source:** | ကျေး ဇူး ပြု ၍ ဖိ နပ် ချွတ် ပါ ။ |
| **Reference:** | ဖိနပ် ချွတ် ကျေးဇူးပြု၍ ။ |
| PBSMT hyp: | **ကျေးဇူးပြု၍ နပ် ဖိ** ချွတ် ။ |
| HPBSMT hyp: | ကျေးဇူးပြု၍ ဖိနပ် ချွတ် ။ |
| OSM hyp: | **ဟော့ကျေးဇူးပြု၍ နပ် ။ ဖိ** ချွတ် |

- Some extra words are found in PBSMT and HPBSMT models for the MSL→MWT translation on both syllable and word segmentations.

| **Syllable - Syllable (sl-my)** | |
|---|---|
| **Source:** | ဆင်း နေ ရာ ဘာ လဲ ပြော ။ |
| **Reference:** | ဘယ် မှာ ဆင်း ရ မ လဲ ပြော ပါ ။ |
| PBSMT hyp: | **ငါ တို့** ဘယ် **နေ ရာ** ဆင်း ရ မ လဲ ပြော ပါ ။ |
| HPBSMT hyp: | ဆင်း **နေ ရာ က ဘာ** လဲ ပြော ပါ ။ |
| OSM hyp: | ဘယ် မှာ ဆင်း ရ မ လဲ ပြော ပါ ။ |
| | |
| **Word-Word(my-sl)** | |
| **Source:** | ကျွန်တော် သောက် စရာ တစ် ခု ခု လို ချင် တယ် ။ |
| **Reference:** | သောက်စရာ တစ်ခုခု လိုချင် ။ |
| PBSMT hyp: | ငါ ခု သောက် စရာ တစ် ခု **ခုံ** ကြိုက် ။ |
| HPBSMT hyp: | ငါ သောက် စရာ တစ် ခု ခု **ခုံ** ကြိုက် ။ |
| OSM hyp: | ငါ ခု သောက် စရာ တစ် ခု ကြိုက် ။ |

# Experiments and Analysis (cont.)
## (NMT Experiments)

### Corpus statistics

- 2,510 Myanmar (my) and MSL (sl) parallel sentences of MSL corpus
- 2,000 sentences were used for training, 310 sentences for development and 200 sentences for evaluation.
- Four types of segmentation schemes, these were word, syllable, SentencePiece and BPE segmentations.

### Moses SMT system

- used the Phrase-based SMT provided by the Moses framework [Kohen et al., 2003]

# Experiments and Analysis (cont.)
## (NMT Experiments)

### Framework for NMT

- used the Sockeye [Hieber et al., 2017] framework, which is based on MXNet [Chen et al., 2015] to train NMT models

### Training Details

- initial learning rate is set to 0.0002,
- size of embeddings and hidden states is 512
- trained all models for maximum epoch using the SGD, Adagrad and Adam optimizers
- size of the training batches were set to 1024, 512, 256, 128 and 64
- byte pair encoding (BPE) models were trained with a vocabulary size of 5,000
- all experiments are run on a single GeForce GTX 1080 8GB ROG STRIX GPU

# Experiments and Analysis (cont.)
## (NMT Experiments)

**BLEU scores of <u>word segmentation</u> for three NMT approaches**

| Source - Target | NMT Approach | Batch Size | Optimizer | BLEU |
|---|---|---|---|---|
| MWT → MSL | RNN | 256 | Adam | 13.10 |
| | CNN | 128 | Adam | **28.80** |
| | Transformer | 256 | Adagrad | 27.91 |
| MSL → MWT | RNN | 256 | Adam | 12.30 |
| | CNN | 256 | Adam | 28.78 |
| | Transformer | 1024 | Adam | **29.38** |

Table 7. BLEU scores of word segmentation for three NMT approaches: RNN, Transformer and CNN for MWT → MSL and MSL → MWT translations tasks

# Experiments and Analysis (cont.)
## (NMT Experiments)

**BLEU scores of syllable segmentation for three NMT approaches**

| Source - Target | NMT Approach | Batch Size | Optimizer | BLEU |
|---|---|---|---|---|
| MWT→MSL | RNN | 256 | Adam | 15.14 |
| | CNN | 256 | Adam | **32.76** |
| | Transformer | 512 | Adagrad | 29.68 |
| MSL→MWT | RNN | 256 | Adam | 12.17 |
| | CNN | 256 | Adam | 35.02 |
| | Transformer | 1024 | Adam | **38.21** |

Table 8. BLEU scores of syllable segmentation for three NMT approaches: RNN, Transformer and CNN for MWT → MSL and MSL → MWT translations tasks

# Experiments and Analysis (cont.)
## (NMT Experiments)

**BLEU scores of SentencePiece segmentation for three NMT approaches**

| Source - Target | NMT Approach | Batch Size | Optimizer | BLEU |
|---|---|---|---|---|
| MWT → MSL | RNN | 256 | Adam | 12.02 |
| | CNN | 1024 | Adam | **22.69** |
| | Transformer | 256 | Adam | 22.13 |
| MSL → MWT | RNN | 256 | Adam | 8.97 |
| | CNN | 256 | Adam | **23.64** |
| | Transformer | 1024 | Adam | 22.13 |

Table 9. BLEU scores of SentencePiece segmentation for three NMT approaches: RNN, Transformer and CNN for MWT → MSL and MSL → MWT translations tasks

# Experiments and Analysis (cont.)
## (NMT Experiments)

### BLEU scores of BPE segmentation for three NMT approaches

| Source - Target | NMT Approach | Batch Size | Optimizer | BLEU |
|---|---|---|---|---|
| MWT → MSL | RNN | 1024 | Adam | 10.56 |
| | CNN | 256 | Adam | 26.91 |
| | Transformer | 256 | Adagrad | **29.39** |
| MSL → MWT | RNN | 256 | Adam | 28.05 |
| | CNN | 512 | Adam | 27.73 |
| | Transformer | 256 | Adagrad | **32.92** |

Table 10. BLEU scores of BPE segmentation for three NMT approaches: RNN, Transformer and CNN for MWT → MSL and MSL → MWT translations tasks

# Experiments and Analysis (cont.)
## (NMT Experiments)

### SMT performances comparison on four types of segmentation units

| Approach | Segmentation | MWT → MSL | MSL → MWT |
|---|---|---|---|
| PBSMT | Word | 29.69 | 32.26 |
| | Syllable | **35.81** | **35.82** |
| | SentencePiece | 28.80 | 26.30 |
| | BPE | 28.20 | 33.14 |

Table 11. BLEU scores of SMT performances comparison on four types of segmentation units for MWT→MSL and MSL→MWT translations tasks

# Experiments and Analysis (cont.)
## (NMT Experiments)

### NMT performances comparaison on four types of segmentation units

| Source-Target | Approach | Segmentation | BLEU |
|---|---|---|---|
| MWT→MSL | CNN | Word | 28.80 |
| | CNN | Syllable | **32.76** |
| | CNN | SentencePiece | 22.69 |
| | Transformer | BPE | 29.39 |

Table 12. BLEU scores of NMT performances comparaison on four types of segmentation units for MWT→MSL translation

| Source-Target | Approach | Segmentation | BLEU |
|---|---|---|---|
| MSL→MWT | CNN | Word | 29.38 |
| | Transformer | Syllable | **38.21** |
| | CNN | SentencePiece | 23.64 |
| | Transformer | BPE | 32.92 |

Table 13. BLEU scores of NMT performances comparaison on four types of segmentation units for MSL → MWT translation

# Experiments and Analysis (cont.)
## (NMT Error Analysis)

- We focus on the performances of three NMT approaches (RNN, CNN, and Transformer)

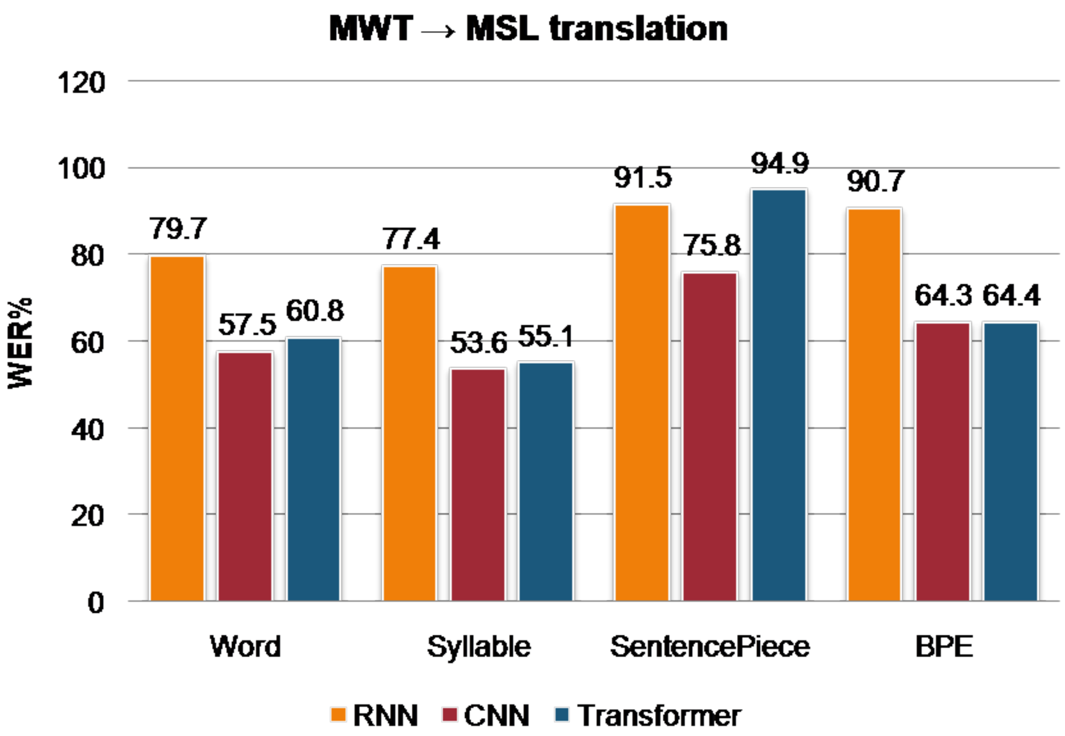



Figure 15: WER of three NMT approaches for MWT to MSL translation for four segmentation schemes

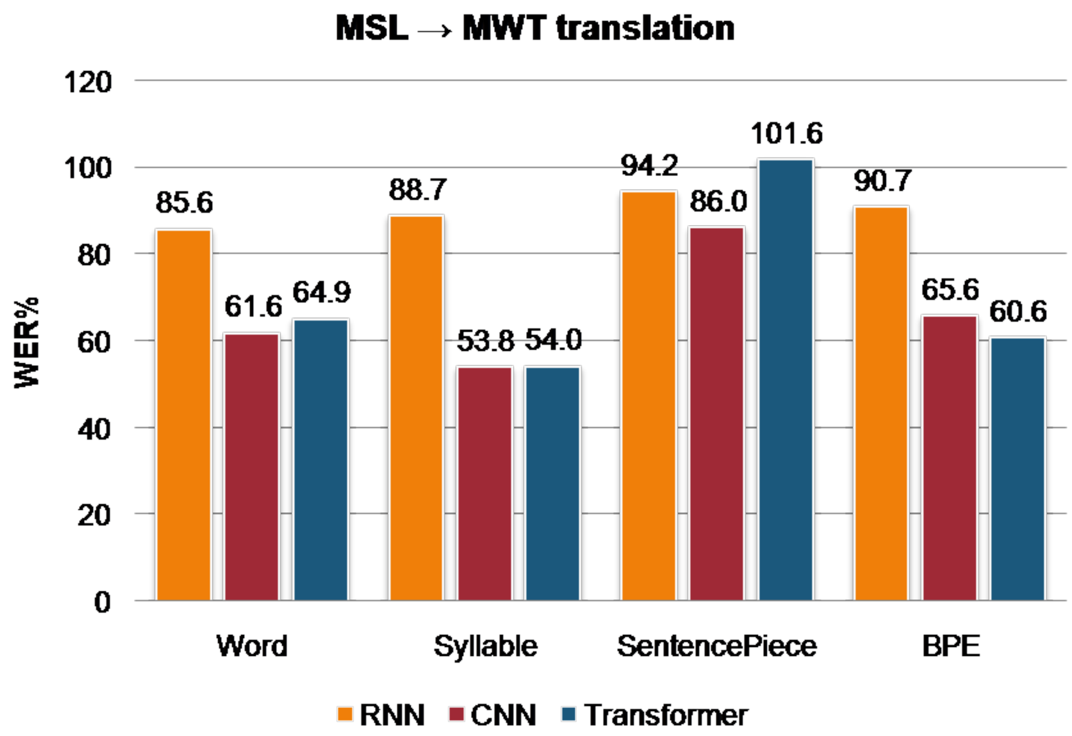



Figure 16: WER of three NMT approaches for MSL to MWT translation for four segmentation schemes

# Experiments and Analysis (cont.)
## (NMT Error Analysis)

The followings are some examples of missing words error (see underline word) that we found on BPE segmentation ("Can I borrow a book?" in English):

**Word Segmentation**

Scores: (#C #S #D #I) 6 0 0 0

REF: စာအုပ် ငှား လို့ ရ မလား ။

HYP: စာအုပ် ငှား လို့ ရ မလား ။

Eval:

**Syllable Segmenation**

Scores: (#C #S #D #I) 8 0 0 0

REF: စာ အုပ် ငှား လို့ ရ မလား ။

HYP: စာ အုပ် ငှား လို့ ရ မလား ။

Eval:

**SentencePiece Segmentation**

Scores: (#C #S #D #I) 5 0 0 1

REF: စာအုပ် ငှား လို့ ရမလ ◌ား ။

HYP: စာအုပ် ငှား လို့ ရမလ ◌ား ။

Eval:

**BPE Segmenation**

Scores: (#C #S #D #I) 5 0 1 0

REF: စာအုပ် ငှား လို့ ရ မလား ။

HYP: <u>****</u> ငှား လို့ ရ မလား ။

Eval: D

# Experiments and Analysis (cont.)
## (NMT Error Analysis)

The followings are some examples of confusion word errors (see underline words) that we found on word, SentencePiece and BPE segmentation ("I am getting stomach pain." in English):

**Word Segmentation**

Scores: (#C #S #D #I) 3 2 0 1

REF: ကျွန်တော် **** ဗိုက် နာ တယ် ။

HYP: ကျွန်တော် ပန်းနာရင်ကျပ် ဖြစ် နေ တယ် ။

Eval: I S S

**SentencePiece Segmentation**

Scores: (#C #S #D #I) 3 2 0 1

REF: ကျွန်တော် **** ဗိုက် နာ တယ် ။

HYP: ကျွန်တော် လည် ချောင် ◌း နာ နေတယ် ။

Eval: I S S

**Syllable Segmenation**

Scores: (#C #S #D #I) 6 0 0 0

REF: ကျွန်တော် ဗိုက် နာ တယ် ။

HYP: ကျွန်တော် ဗိုက် နာ တယ် ။

Eval:

**BPE Segmenation**

Scores: (#C #S #D #I) 6 0 0 0

REF: ကျွန်တော် ဗိုက် နာ တယ် ။

HYP: ကျွန်တော် မျက်စ◌ိ နာ တယ် ။

Eval: S

# Experiments and Analysis (cont.)
## (Unsupervised NMT Experiments)

### Corpus statistics

- For unsupervised NMT experiments, only 4,500 MSL sentences training data were used for all unsupervised NMT models, since MSL data are scarce and no more data to incorporate.

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

Table 14. Four types of Myanmar language monolingual training data

<u>Parallel data</u>

- 650 sentences for development
- 590 sentences for evaluation

# Experiments and Analysis (cont.)
## (Unsupervised NMT Experiments)

### Framework for Unsupervised NMT

- used the UnsupervisedMT [Lample et al., 2018] to train Unsupervised NMT

### Training Details

- Transformer network architecture
- initial learning rate is set to 0.0001
- size of embeddings and hidden states is 512
- trained all models for maximum epoch 1500 using the Adam optimizers
- size of the training batch is set to 32
- byte pair encoding (BPE) models were trained with a vocabulary size of 4,500
- all experiments are run on two GeForce GTX 1080 8GB ROG STRIX GPUs

# Experiments and Analysis (cont.)
## (Unsupervised NMT Experiments)

**BLEU scores of SMT**

| Source - Target | PBSMT | HPBSMT | OSM |
|---|---|---|---|
| MWT→MSL | 39.29 | 40.59 | **41.12** |
| MSL→ MWT | 36.18 | 35.49 | **37.56** |

Table 15. BLEU scores of supervised NMT

**BLEU scores of supervised NMT**

| Source - Target | Epoch 500 | Epoch 1000 | Epoch 1500 |
|---|---|---|---|
| MWT→MSL | 16.04 | 20.46 | **23.92** |
| MSL→ MWT | 19.46 | 22.89 | **25.28** |

Table 16. BLEU scores of supervised NMT

# Experiments and Analysis (cont.)
## (Unsupervised NMT Experiments)

**BLEU scores unsupervised NMT for Training data 1 (MSL data only)**

| Source - Target | Epoch 500 | Epoch 1000 | Epoch 1500 |
|---|---|---|---|
| MWT→MSL | **28.99** (22.40) | 28.84 (23.75) | 28.74 **(24.01)** |
| MSL→ MWT | **15.04** (11.58) | 13.10 **(12.03)** | 12.52 (11.75) |

Table 17. BLEU scores of training data 2 for MWT→MSL and MSL→MWT translations tasks

**BLEU scores of unsupervised NMT for Training data 2 (MSL and Primary English)**

| **Source - Target** | Epoch 500 | Epoch 1000 | Epoch 1500 |
|---|---|---|---|
| MWT→MSL | 27.79 (27.04) | **29.06** **(27.56)** | 27.69 (27.21) |
| MSL→ MWT | **11.21** **(10.47)** | 9.76 (9.04) | 9.61 (8.08) |

Table 18. BLEU scores of training data 2 for MWT→MSL and MSL→MWT translations tasks

# Experiments and Analysis (cont.)
## (Unsupervised NMT Experiments)

**BLEU scores of <u>unsupervised NMT</u> for Training data 3 (MSL and myPOS)**

| Source - Target | Epoch 500 | Epoch 1000 | Epoch 1500 |
|---|---|---|---|
| MWT→MSL | 28.98<br>(28.30) | **30.13**<br>(29.40) | 30.04<br>**(29.53)** |
| MSL→ MWT | **12.06**<br>**(9.98)** | 9.88<br>(9.04) | 9.51<br>(8.72) |

Table 19. BLEU scores of training data 3 for MWT→MSL and MSL→MWT translations tasks

**BLEU scores of <u>unsupervised NMT</u> for Training data 4 (MSL, Primary English and myPOS)**

| Source - Target | Epoch 500 | Epoch 1000 | Epoch 1500 |
|---|---|---|---|
| MWT→MSL | 26.75<br>(24.33) | 28.19<br>(27.03) | **28.88**<br>**(27.52)** |
| MSL→ MWT | 6.34<br>(5.76) | 7.44<br>(6.81) | **8.44**<br>**(6.88)** |

Table 20. BLEU scores of training data 4 for MWT→MSL and MSL→MWT translations tasks

# Experiments and Analysis (cont.)
## (Unsupervised NMT Error Analysis)

- We focus on the performances of unsupervised NMT approach

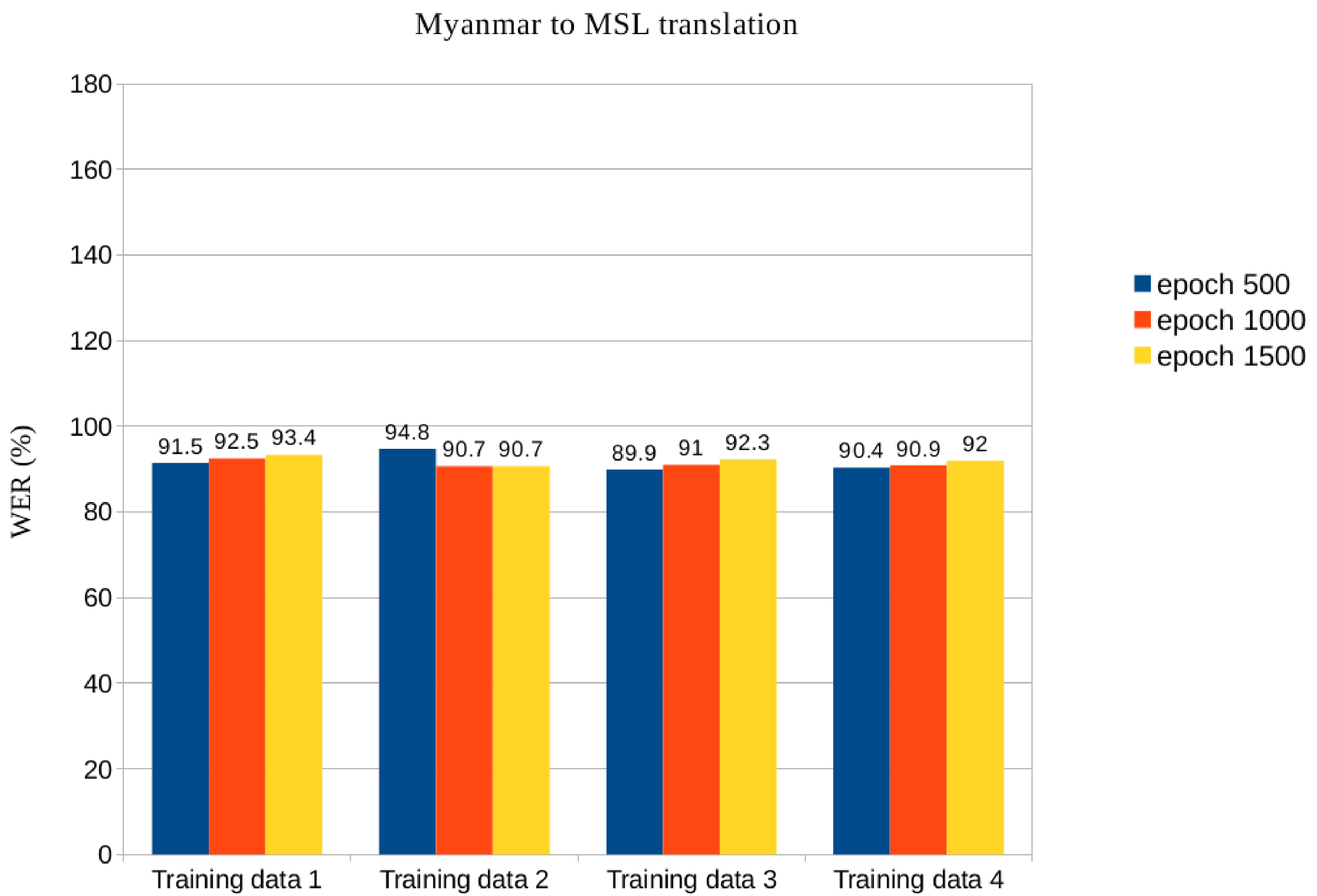



Figure 17. WER of unsupervised NMT approach for Myanmar to MSL translation

# Experiments and Analysis (cont.)
## (Unsupervised NMT Error Analysis)

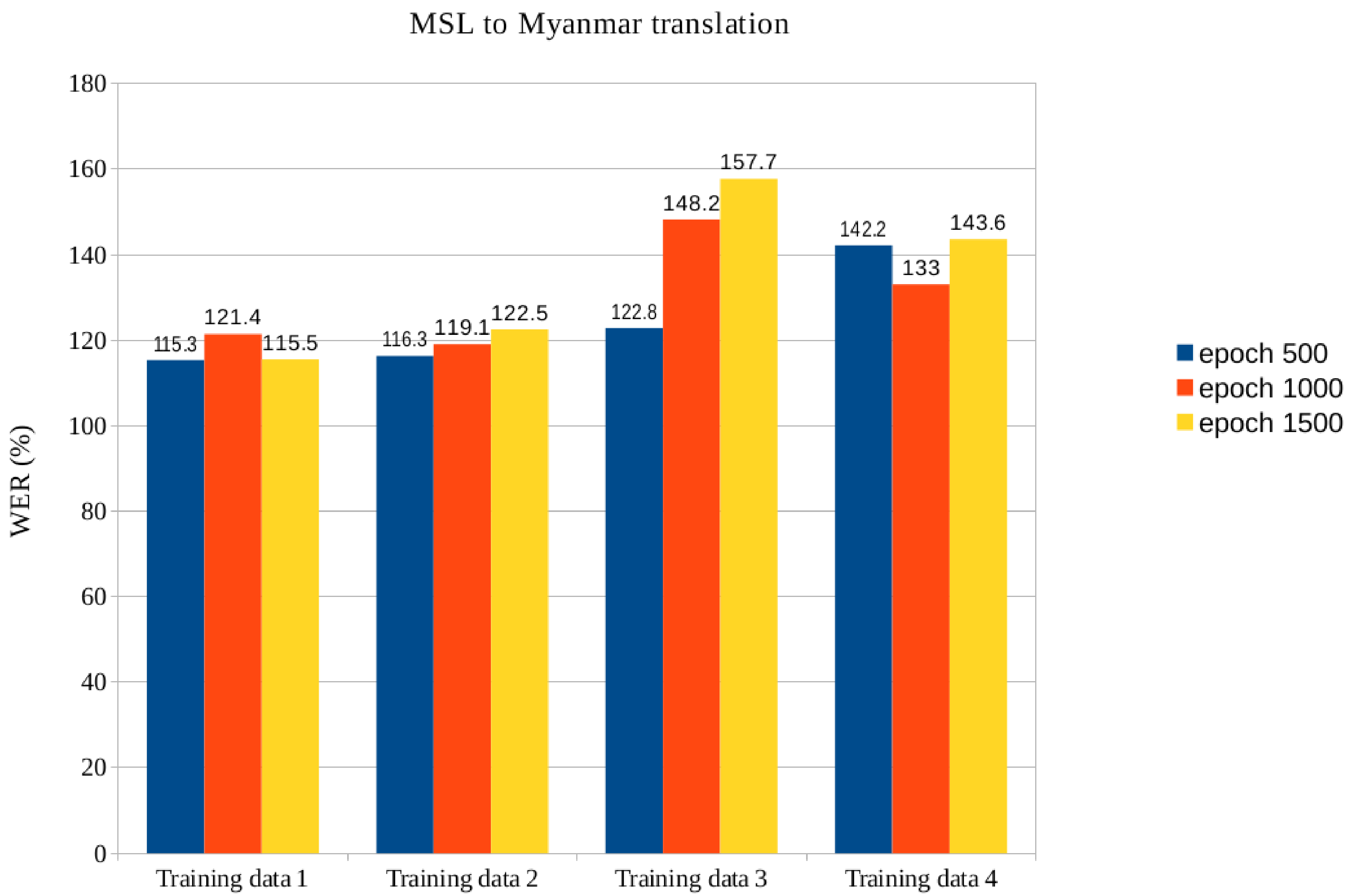



Figure 18. WER of unsupervised NMT approach for MSL to Myanmar translation

# Experiments and Analysis (cont.)
## (Unsupervised NMT Error Analysis)

Detail analysis on confusion pairs of unsupervised NMT approache, most of the confusion pairs are caused by the three 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 corpus

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

Table 21. Top 10 confusion pairs of unsupervised NMT model for Myanmar to MSL translation

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

Table 22. Top 10 confusion pairs of unsupervised NMT model for MSL to Myanmar translation

# Conclusions and Future Work

- This proposed system has presented the first study of the SMT, NMT and Unsupervised NMT between Myanmar sign language and Myanmar written text.
- Implemented three SMT systems (PBSMT, HPBSMT and OSM), three NMT systems (RNN, CNN, Transformer) and Unsupervised NMT system with developing MSL-MWT corpus.
- The proposed system investigated the effectiveness of four word embeddings schemes (syllable, word (sign unit-based) , SentencePiece and BPE) .
- For SMT systems, HPBSMT and OSM with syllable segmentation for both source and target languages achieved the highest BLEU scores for translation between MSL and MWT.
- For NMT systems, Transformer outperformed both CNN and RNN for MWT-to-MSL and MSL-to-MWT translation tasks.
- Highest quality NMT and SMT performances were attained with syllable segmentation for both MSL and MWT.
- For Unsupervised NMT system, training data (MSL-MWT and myPOS corpora) gives highest BLUE scores for MWT to MSL translation task.
- We planning our research work for automatic MSL video to text caption in the near future.

# References


https://www.cs.upc.edu/~cristinae/CV/docs/tutorialSMTprint.pdf

https://engineering.fb.com/ml-applications/a-novel-approach-to-neural-machine-translation/

https://yashuseth.blog/2019/03/03/how-can-unsupervised-neural-machine-translation-work/

https://engineering.fb.com/ai-research/unsupervised-machine-translation-a-novel-approach-to-provide-fast-accurate-translations-for-more-languages/

https://iconictranslation.com/2018/09/issue-11-unsupervised-neural-mt/

https://jalammar.github.io/illustrated-transformer/

https://jalammar.github.io/visualizing-neural-machine-translation-mechanics-of-seq2seq-models-with-attention/

Convolutional Sequence to Sequence Learning. Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, Yann N. Dauphin. ArXiv, 2017

SOCKEYE:A Toolkit for Neural Machine Translation. Felix Hieber, Tobias Domhan, Michael Denkowski,David Vilar, Artem Sokolov, Ann Clifton, Matt Post. ArXiv, 2018

http://u.cs.biu.ac.il/~yogo/courses/mt2013/l5/11-koehn-phrase-extraction.pdf

# Publication List

1. **Swe Zin Moe**, Ye Kyaw Thu, Hlaing Myat Nwe, Hnin Wai Wai Hlaing, Ni Htwe Aung, Khaing Hsu Wai, Hnin Aye Thant and Nandar Win Min, "Development of Natural Language Processing based Communication and Educational Assisted Systems for the People with Hearing Disability in Myanmar", In the Journal "LINGUAS & LETRAS", volume 20, n. 48, Western Paraná State University, State of Paraná, Brazil. [on going process]

2. **Swe Zin Moe**, Ye Kyaw Thu, Hnin Aye Thant and Nandar Win Min, "Unsupervised Neural Machine Translation between Myanmar Sign Language and Myanmar Language", In the Journal of Intelligent Informatics and Smart Technology (JIIST), Thailand. [on going process]

3. **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.

4. **Swe Zin Moe**, Ye Kyaw Thu, Hnin Wai Wai Hlaing, Hlaing Myat Nwe, Ni Htwe Aung, Hnin Aye Thant, Nandar Win Min, "Statistical Machine Translation between Myanmar Sign Language and Myanmar Writen Text", In Proceedings of ICCA2018, February 22-23, 2018, Yangon, Myanmar, pp. 217-227.

# Publication List (con't)

5. Hnin Wai Wai Hlaing, Ye Kyaw Thu, **Swe Zin Moe**, Hlaing Myat Nwe, Ni Htwe Aung, Nandar Win Min, Hnin Aye Thant, "Statistical Machine Translation between Myanmar Sign Language and Myanmar SignWriting", In Proceedings of the 1st International Symposium on Artificial Intelligence for ASEAN Development (ASEAN-AI 2018), March 26-27, 2018, Phuket, Thailand, pp. 65-72.

6. Hlaing Myat Nwe, Ye Kyaw Thu, Hnin Wai Wai Hlaing, **Swe Zin Moe**, Ni Htwe Aung , Hnin Aye Thant , Nanda Win Min, "Two Fingerspelling Keyboard Layouts for Myanmar SignWriting", In Proceedings of ICCA2018, February 22-23, 2018, Yangon, Myanmar, pp. 290-298.

7. Khaing Hsu Wai, Ye Kyaw Thu, Hnin Aye Thant, **Swe Zin Moe** and Thepchai Supnithi, "String Similarity Measures for Myanmar Language (Burmese)", The First Workshop on NLP Solutions for Under Resourced Languages (NSURL 2019), 11-13 September 2019, Trento, Italy

# Acknowledgement


- 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.

**Thank you**

**for your attention!**

Thank you
for your attention!