Machine Translation Summit 2023

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

Hlaing Myat Nwe, Kiyoaki Shirai, Natthawut Kertkeidkachorn, Thanaruk Theeramunkong, Ye Kyaw Thu, Thepchai Supnithi, Natsuda Kaothanthong

Email: hlaingmyatnwe@jaist.ac.jp

7 SEP 2023

Outline

- Background
- Goal
- Related Work
- Proposed Method
- Evaluation
- Conclusion
- References

Background

- There are many deaf or hard-of-hearing people in Myanmar
  - Approximately 1.1 M of the population
  - Population of deaf people with a university education is only 0.0006%
- They mostly rely on Myanmar Sign Language (MSL) for communication
- However, both deaf and hearing people find it rather difficult to learn MSL
- There are few and limited assistive technology available for deaf people

Goal

- To develop a high-quality translation system between Myanmar Sign Language (MSL) and Myanmar Written Language (MWL)
- Technical challenge
  - There are only a few limited parallel corpora between MSL and MWL
- Two techniques are used:
  - Transfer learning
  - Semi-supervised learning with self-training

Related Work (1/2)

1. Machine Translation for Low-Resource Languages Using Transfer Learning

- **Zoph et al. (2016)** [7] proposed transfer learning for low-resource neural machine translation (NMT). Their method involves training the parent model with a high-resource language pair and transferring some information from it to the child model with a low-resource language pair. The parent model's parameters influence the child model, boosting baseline NMT performance by an average of 5.6 BLEU.

- **Kocmi and Bojar (2018)** [9] proposed an innovative transfer learning approach for NMT. Parent model from any language pair improves child model performance, even with unrelated languages and different alphabets.

- **Maimaiti et al. (2022)** [10] proposed language-independent Hybrid Transfer Learning (HTL), which enhances NMT quality for low-resource, morphologically rich languages. HTL shares lexicon embeddings between parent and child languages, outperforming five state-of-the-art methods for Azerbaijani and Uzbek translations.

Related Work (2/2)

2. Myanmar Sign Language (MSL) Translation

- There are only a few previous studies since limited parallel data availability makes MSL translation difficult

- **Moe et al. (2018a)** [11] explored statistical machine translation (SMT) for MSL to MWL translation. Optimal performance was achieved using the Operation Sequence Model and Hierarchical Phrase-based SMT with syllable-based segmentation.

- **Moe et al. (2018b)** [12] compared NMT approaches for MSL translation. The transformer model outperformed Convolutional and Recurrent Neural Networks.

- **Moe et al. (2020)** [13] investigated unsupervised neural machine translation (U-NMT) for MSL and MWL. Among several monolingual corpora, the highest BLEU score is obtained when the myPOS corpus is used.

Characteristics of this study

- Transfer learning from a parent MT model (of high-resource language) to a child model (of low-resource language) is applied to the translation of MSL
  - Although transfer learning is common, it has not been applied to MSL

- In addition, semi-supervised learning by self-training is applied to increase the size of the parallel corpus between MSL and MWL

Myanmar Sign Language (MSL) (1/2)

- Deaf people use MSL as their primary communication language instead of voice
- There are two types of features: manual features (MF) and non-manual features (NMF) to convey meaning

Manual Features (MF)

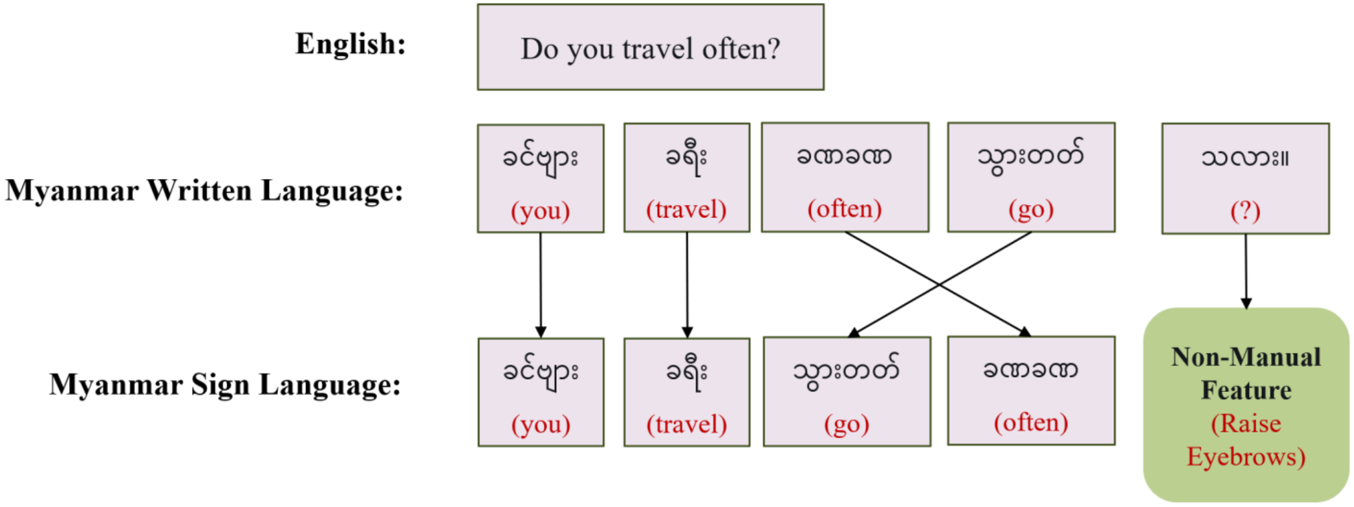
Hand shape, movement, position

Non-manual Features (NMF)

eye, mouth, chin movement, facial expression, & so on

Myanmar Sign Language (MSL) (2/2)

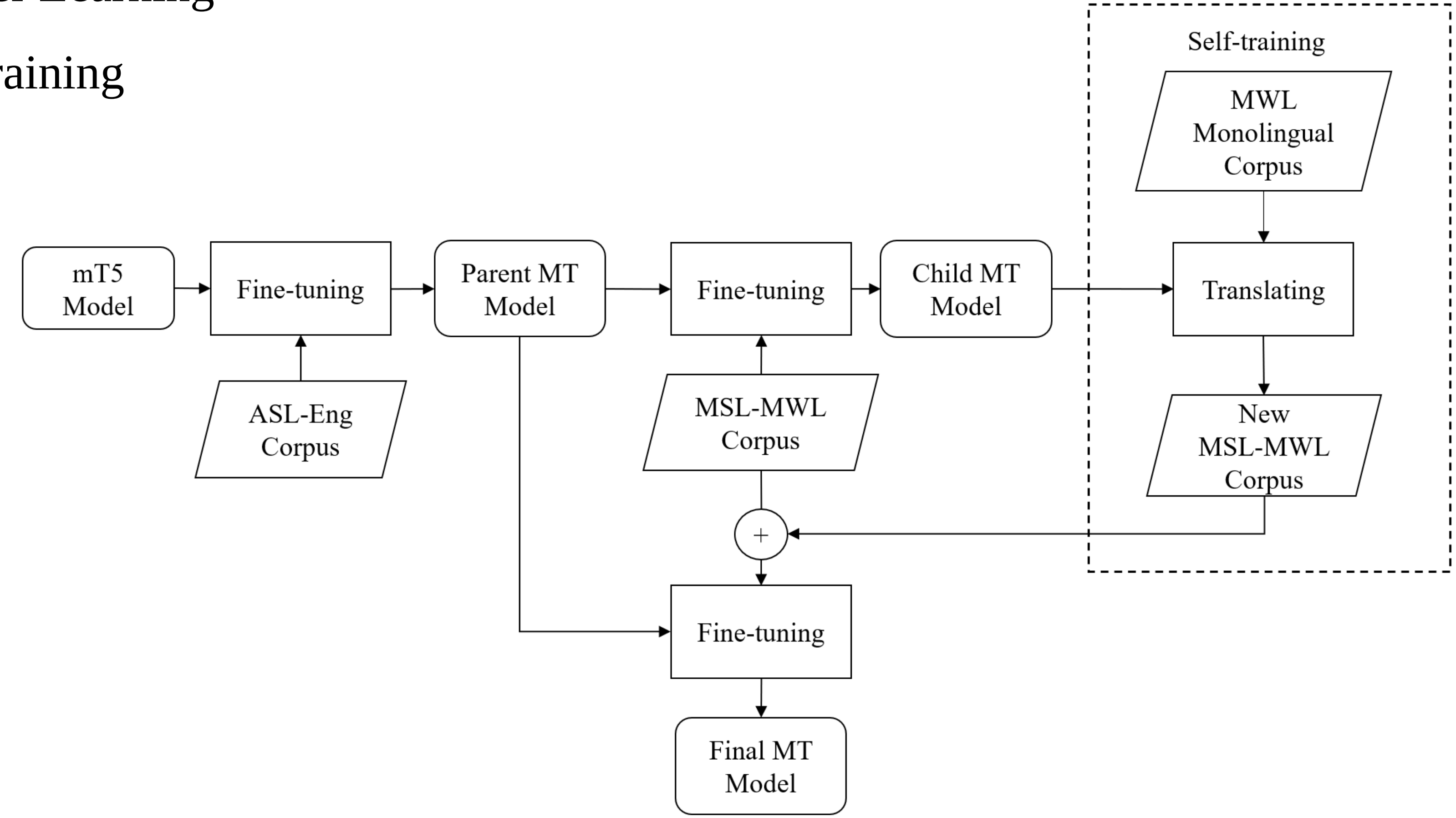
- MWL is tonal and syllable-based, whereas MSL relies on visual-spatial elements to convey meaning
- MSL and MWL are different in grammatical structure



- This study focuses on translating word sequences from MSL to MWL and vice versa, where a word in MSL represents a sign
- Currently, our approach can convert an MSL gloss into a sign or gesture, it doesn't consider for non-manual features

Proposed Method

- Multilingual Pre-trained Text-to-Text Transfer Transformer (mT5) model is used as a based translation system
- Our method consists of two basic methodologies
  1. Transfer Learning
  2. Self-Training



Preprocessing

- Sentences in MSL and MWL are segmented into a sequence of tokens as a preprocessing step
- Three segmentation schemes are used:
  - **Word-based Segmentation**: A sentence is manually divided into a word sequence by using spaces
  - **Syllable-based Segmentation**: The regular expression (RE)-based Myanmar syllable segmentation tool, called **sylbreak**, is used
  - **Byte-pair Encoding (BPE)-based Segmentation**: The BPE-segmentation model is built from the large monolingual corpus of MWL, myPOS

	MWL	MSL
No Segmentation:	ကျေးဇူးပြု၍ဖိနပ်ချွတ်ပါ။ (Please take off your shoes.)	ဖိနပ်ချွတ်ပါကျေးဇူးပြု၍။ (Please take off your shoes.)
Word-based Segmentation:	ကျေးဇူးပြု၍ (Please) ဖိနပ် (Shoes) ချွတ် ပါ (Take off) ။	ဖိနပ် (Shoes) ချွတ်ပါ (Take off) ကျေးဇူးပြု၍ (Please) ။
Syllable-based Segmentation:	ကျေး ဇူး ပြု ၍ ဖိ နပ် ချွတ် ပါ ။	ဖိ နပ် ချွတ် ပါ ကျေး ဇူး ပြု ၍ ။
BPE-based Segmentation:	ကျေးဇူးပြု၍ ဖိနပ် ချွတ် ပါ ။	ဖိနပ် ချွတ်@@ ပါ ကျေးဇူးပြု၍ ။

Transfer Learning

- The pre-trained mT5-Base model, which has 580 million parameters, is used

- Two-step fine-tuning of the mT5 model:
  - The parallel corpus of American Sign Language (ASL) and English is used to build the parent MT model
  - The parent MT model is fine-tuned again using a parallel corpus of 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 a Myanmar parallel corpus

Self-Training

- The sentences in the monolingual corpus are translated into MSL sentences using the trained child MT model
- The pairs of the original monolingual sentences and the translated sentences form a new parallel corpus
- To improve the quality of the automatically constructed parallel corpus, we filtered out the unreliable translations
  - The score of the translated sentence *s* is calculated using the following equation:

$$score\ (s) = \log P(s) \cong \sum_{w_i \in s} \log P_{mT5}(w_i)$$

  - $P(s)$ is the probability of generating s, while $P_{mT5}(w_i)$ is the probability of generating i-th token w_i estimated by the fine-tuned mT5 model
  - The top *N* translations with the highest scores are kept to make the new parallel corpus
- Finally, the original and new parallel corpus is used to fine-tune the child MT model

Evaluation (1/2)

Dataset

i. **Parallel Corpus of English:** A part of English-ASL Gloss Parallel Corpus2012 (ASLG-PC12) is used for training the parent MT model

ii. **Parallel Corpus of Myanmar Language:** Only one parallel corpus of MSL and MWL, which was collected from 30 sign language trainers and deaf people

iii. **Monolingual Corpus of Myanmar Language:** The myPOS corpus consists of 43,196 sentences that have been manually word-segmented and POS-tagged for the purpose of NLP research and development

	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.

Evaluation (2/2)

1. **Bilingual Evaluation Understudy (BLEU) score**
   - To compare MT systems using different segmentation schemes, it is measured by counting the overlap of character n-gram

2. **Word Error Rate (WER)**

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

S = the number of substitution errors

D = the number of deletion errors

I = the number of insertion errors

N = the total number of tokens in the reference

Result and Discussion (1/4)

- Several MT models between MSL and MWL are trained and compared
  - Three segmentation schemes (word, syllable, BPE)
  - The models trained with and without transfer learning
  - The models trained with and without the enlarged parallel corpus obtained by self-training

- The BLEU scores are calculated with confidence interval values at the **significance level of 0.95**

Result and Discussion (2/4)

- Comparing the models mT5 and mT5+T, an improvement of **0.62 points in MSL → MWL** and **5.06 points in MWL → MSL** with the syllable-based segmentation
- Comparing the models mT5 and mT5+S, the maximum improvement is **4.7 points of the MT model for MWL →MSL** with the syllable segmentation
- Combining transfer learning and self-training can further boost MT performance, the highest BLEU scores are **56.60 and 57.11 for MSL → MWL and MWL → MSL**, which are 5.93 and 5.88 points higher than the baseline model

BLEU score (↑)

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

Result and Discussion (3/4)

- The lowest WER is obtained by mT5+T+S with the syllable segmentation scheme
- 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
  3. The contributions of those two techniques are comparable

WER (%) (↓)

Model	Parent Model	Self-Training	MSL→MWL word	MSL→MWL syllable	MSL→MWL BPE	MWL→MSL word	MWL→MSL syllable	MWL→MSL 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	$\boxed{\textbf{48.2}}$	50.4	55.2	$\boxed{\textbf{46.5}}$	49.2

Result and Discussion (4/4)

- Comparison between our best method (mT5+T+S, syllable) and three previous studies
- Syllable-based BLEU score is measured due to the consistency with previous work
- Our method is better than or comparable to the previous studies
  - It is not a fair comparison since the size of the datasets used for the evaluation are different

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

Error Analysis

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?

- Investigating the errors of the best model mT5+T+S for translating from MSL to MWL with the syllable segmentation scheme

- Example E1 shows that the model generates the word **"they"** as the first word, despite it being the sixth word in the reference
- This highlights the model's struggle in capturing grammatical structure differences between MSL and MWL

- In example E2, the word **"she"** in the reference is replaced with **"he"**, causing a gender inconsistency with "girl"

- 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
- This indicates that some substitution errors are acceptable, even if the output differs from the reference

Conclusion

Summary

- A novel method is proposed 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, transfer learning and self-training were applied where mT5 pre-trained model was used as the backbone

- The results of the experiments showed that both transfer learning and self-training could contribute to improving MT performance

Future Work

- We will extend our method for the translation of MSL to include both manual and non-manual features

References (1/2)

1. Sylbreak tool: https://github.com/ye-kyaw-thu/sylbreak
2. Subword Neural Machine Translation (subword-nmt) library: https://github.com/rsennrich/subword-nmt
3. aslg-pc12 dataset: https://huggingface.co/datasets/aslg
4. myPOS corpus: https://github.com/ye-kyaw-thu/myPOS
5. The bleukit-NTCIR7 Scoring tool: http://www.nlp.mibel.cs.tsukuba.ac.jp/bleu_kit/
6. The SCLITE (Score Lite) program: https://github.com/usnistgov/SCTK
7. Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 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, November 2016. Association for Computational Linguistics.
8. Raj Dabre, Tetsuji Nakagawa, and Hideto Kazawa. 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), November 2017.
9. Tom Kocmi and Ondřej Bojar. 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, October 2018.
10. Mieradilijiang Maimaiti, Yang Liu, Huanbo Luan, and Maosong Sun. Enriching the transfer learning with pre-trained lexicon embedding for low-resource neural machine translation. Tsinghua Science and Technology, 27(1):150–163, 2022.

References (2/2)

11. Swe Zin Moe, Ye Kyaw Thu, Hnin Wai Wai Hlaing, Hlaing Myat Nwe, Ni Htwe Aung, Hnin Aye Thant, and Nandar Win Min. Statistical machine translation between Myanmar sign language and Myanmar written text. In 16th International Conference on Computer Applications, Yangon, Myanmar, February 22-23, 2018.
12. 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), Phnom Penh, Cambodia, December 13-14, 2018.
13. Swe Zin Moe, Ye Kyaw Thu, Hnin Aye Thant, Nandar Win Min, and Thepchai Supnithi. Unsupervised neural machine translation between Myanmar sign language and Myanmar language. Journal of Intelligent Informatics and Smart Technology, 1(1):53–61, April 2020.
14. Win Pa Pa, Ye Kyaw Thu, Andrew Finch, and Eiichiro Sumita. 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, 26-28 August 2015. Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research, 21(140):1–67, 2020.
15. Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 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, June 2021.

Thank You!

Your time and attention are greatly appreciated.

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

Presented by

Hlaing Myat Nwe

(Email: hlaingmyatnwe@jaist.ac.jp)