

ASEAN Language Speech Translation thru' U-STAR

Dr. LI Haizhou, NUS, Singapore

Ms. Aw Ai Ti, I2R, Singapore

Speaker: Tarun Kumar Vangani, I2R, Singapore

Background :

Speech-to-speech translation system, which allows a user's speech to be translated into designated languages with synthesized voices, is a challenging and ongoing worldwide research carried out in many research institutes.

Such systems are expected to be the nexus of services delivery and applications for consumers and businesses in the future to bridge the language gaps.

Targets:

- Development of a service (delivered as a mobile application) to provide speech-to-speech translation leveraging U-STAR infrastructure.
- Development of language resources to facilitate the research & development of ASEAN speech translation technologies

Project Members :

Name	Organization	Country
Dr. Rapid Sun	Director of Research and Development Center, NIPTICT	Cambodia
Dr. Hammam Riza	Deputy Chairman IT, Energy and Material, BPPT	Indonesia
Prof. Sevia M. Idrus	Professor, Communication Engineering Department, Faculty Of Electrical Engineering, UTM	Malaysia
Prof. Khin Mar Soe	Professor, NLP Lab, UCSY	Myanmar
Prof. Li Haizhou	Professor at the Department of Electrical and Computer Engineering, and the Department of Mechanical Engineering of the National University of Singapore.	Singapore
Ms Aw Ai Ti	Head, Aural & Language Intelligence Department, I2R, Singapore	Singapore
Dr. Chai Wutiwivatchai	Executive Director, National Electronics and Computer Technology Center (NECTEC)	Thailand
Prof. NGUYEN Thi Thu Trang	Assistant Professor in Department of Software Engineering, School of Information and Communication Technology, HUST	Vietnam
Prof. Luong Chi Mai	Assoc. Prof, Speech and Language Processing, IOIT	Vietnam

Project Duration : Started on **1 July 2016** and Completed on **30 June 2019**.

- uniTRANS - a mobile based speech-to-speech translation application was developed to facilitate communication among ASEAN community.
- The interface was localized to native languages of project team members to encourage use amongst speaker of different languages.



- **Data Specifications**

- Content of utterances: should be common and appropriate for general use, and reflecting a variety of practical situations.
- Target Speakers: as many speakers as possible and well distributed across the ages 15-75 years (minimum of 5 male speakers and 5 female speakers)
- Speaking style and environment: fluent expressions, and these should be read naturally and fluently by the speakers (no fillers, hesitations, and long pauses are recorded)

- **Transcription Specifications**

- Transcription of Numbers, Acronyms, Foreign words and names

- **Method of Data Collection**

- Recording Device: Device and mechanism used for data collection
- Speech Data Files and Naming Conventions

- **Translation Specifications**

- Punctuation Insertion & Translation of Number: as natural as possible (e..g phone number:12345678)
- Translation or Transliteration: always translate unless no English equivalent

- I2R collaborated with UCSY on two projects through student attachment:
 - Myanmar-to-English translation using syllable-based neural machine translation technique (Yi Mon Shwe Sin, Khin Mar Soe, UCSY; Wu Kui, Aw Ai Ti , I2R)
 - Myanmar word segmentation (Hsu Myat Mo , Khin Mar Soe, UCSY; Zhou Nina, Aw Ai Ti, I2R)
- Two workshops were conducted to facilitate the sharing of knowledge among the team members.
 - First Workshop: 4 October 2016
 - Second Workshop: 6 December 2017.



Language	Organization	Country	Data Statistics
Khmer	NIPTICT	Cambodia	<ul style="list-style-type: none"> • 10K utterances collected and translated. • 4K utterances selected to record as voice data.
Bahasa Indonesia	BPPT	Indonesia	<ul style="list-style-type: none"> • 5000 utterances collected, translated and recorded.
Bahasa Melayu	UTM	Malaysia	<ul style="list-style-type: none"> • 5000 utterances collected, translated and recorded.
Myanmar	UCSY	Myanmar	<ul style="list-style-type: none"> • 4000 utterances collected and translated.
Chinese	I2R	Singapore	<ul style="list-style-type: none"> • 5000 utterances collected, translated and recorded.
Vietnamese	HUST	Vietnam	<ul style="list-style-type: none"> • 6,500 Vietnamese text utterances collected • 3,000 Parallel text utterances • 1,200 recorded utterances (Vietnamese)
Vietnamese	IOIT	Vietnam	<ul style="list-style-type: none"> • 2000 utterances collected and recorded
Thai	NECTEC	Thailand	<ul style="list-style-type: none"> • 6000 utterances have been collected, translated, and NE annotated • 4000 utterances recorded

The presentations at International Conference:

No:	Paper title:	Author names	Affiliation	Conference name:	The date of the conference	The venue of the conference
1.	A Hybrid Method for Vietnamese Text Normalization	<p>NGUYEN Thi Thu Trang</p> <p>DANG Xuan Bach</p> <p>NGUYEN Xuan Tung</p>	<p>VBee Group Hanoi University of Science and Technology Hanoi, Vietnam trangntt@soic.t.hust.edu.vn</p> <p>Hanoi University of Science and Technology Hanoi, Vietnam bachdx96@gmail.com</p> <p>VBee Group Hanoi University of Science and Technology Hanoi, Vietnam tungxbk@gmail.com</p>	<p>3rd International Conference on Natural Language Processing and Information Retrieval</p>	<p>June 28 - 30, 2019</p>	<p>Tokushima, Japan</p>

- **Background**

- Myanmar scripts are written continuously as a sequence of characters without any delimiter between words.
- Word is formed by characters and syllables with certain rules.

- **Method**

- Existing methods use dictionary, rule, traditional machine learning approaches e.g., HMM or CRF.
- Investigate syllable-based segmentation scheme and use Bi-LSTM for sequential labelling

Unsegmented sentence:

အတိတ်မှသင်ခန်းစာသည်အနာဂတ်ရဲ့အားအင်။

Syllable-based segmented sentence:

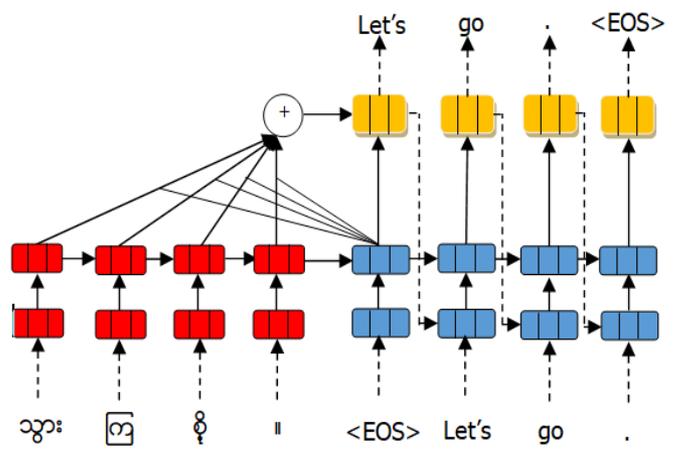
အ/Bတိတ်/Iမှ/Bသင်/Bခန်း/Iစာ/Iသည်/Bအ/Bနာ/Iဂတ်/Iရဲ့/Bအား/Bအင်/I။/B

#	Sent.	Word	Syllable	Char
Train	47K	1.0M	1.6M	4.785M
Dev.	2K	46,500	59,296	151,447
Test	2K	38,000	60,004	226,583

CRF	Prec	Recall	F 1
Char-based	94.03	93.93	93.98
Syllable-based	93.9	94.08	93.99

Models		Precision/recall/F-measure	
		Dev.	Test
BiLSTM (SGD)	Char	89.40/ 88.94 / 88.48	81.67/ 84.13 / 82.88
	Syllable	93.67 / 94.0 / 93.83	92.73/ 92.40 / 92.56
BiLSTM-CRF (SGD)	Char	88.23/ 87.79 / 87.35	82.32/ 83.04 / 82.68
	Syllable	92.03/ 93.59 / 92.80	91.20/ 91.90 / 91.55

- Word level neural machine translation (NMT) cannot model rare words in translating languages with rich morphology. Therefore, a syllable-based NMT is presented by segmenting the source sentences at syllable level using a Myanmar syllable segmenter.
- Syllable-based NMT cannot translate unknown words well and a Name Entity database is incorporated to improve name entity translation.



Myanmar-English	Parallel Sentences
Original Corpus	228,767
NE-database	27,024
Total	255,791

System	BLEU
Word-based NMT	21.88
Character-based NMT	20.71
Syllable-based NMT	26.50
Syllable-based NMT+ NE-database	27.94

- A hybrid method for normalizing written text often found on newspapers to its spoken form.
- The first step involves classifying Non-Standard-Words into different categories using Random Forest. The latter one is to expand them, depending on their types, into pronounceable syllables using a hybrid method.
- The classification model gives an enhanced result of 99.20% with the category combination and the feature optimization. In the expansion, the sequence-to-sequence model shows a good result of 96.53% for abbreviations and 96.25% for loanwords with a post-adjustment for some completely wrong cases.

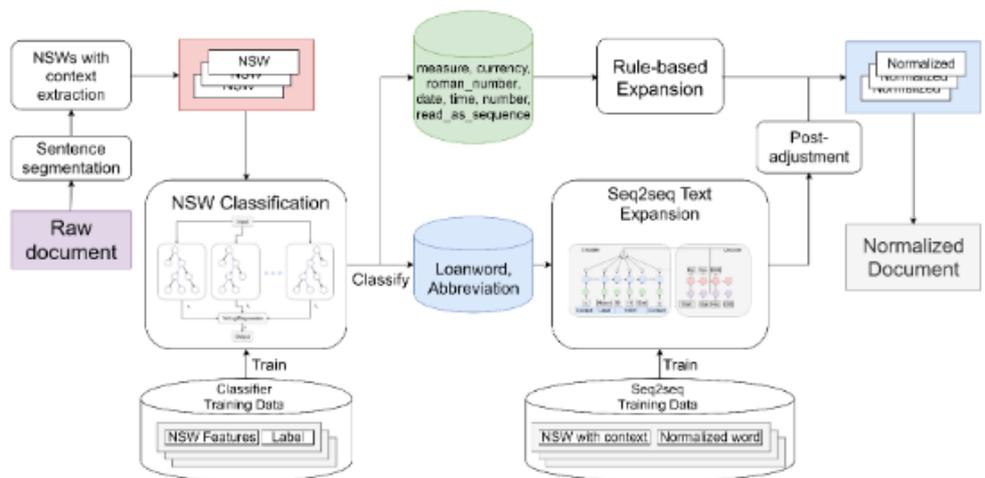


Figure 2. Proposed architecture for text normalization.

- Data Collection and standardization of data policy for speech-to-speech translation.
- Development of a Speech-to-Speech system made publicly available.
- Development of a Myanmar-to-English translation system prototype.

- Leveraging U-STAR infrastructure to collaborate on speech translation technologies.
- Development of innovative applications using U-STAR services to benefit the community.
- Collaboration on building up data resources.

