English to Burmese Machine Translation with Asian Pivot Languages

Wint Theingi Zaw, Ye Kyaw Thu, Zar Zar Hlaing, Thepchai Supnithi

Abstract— **A parallel corpus plays an important role in both statistical and neural machine translations. There are several parallel corpora for resource-rich language pairs. However, parallel corpora are usually readily unavailable for low-resource language pairs such as Chinese-Burmese, English-Burmese, and Thai-Burmese. Bridging source and target languages with a pivot language is an important approach and it has been especially widely used in statistical machine translation. This paper investigates the appropriateness of Asian languages as pivot languages for English-to-Burmese machine translation with freely available online machine translation services. The five small English-Burmese parallel corpora were constructed based on the five English-Burmese bilingual books for the evaluation process. The experimental results with Yandex, Baidu, Bing, and Google machine translation services show that the pivot machine translation approach without using parallel corpora provide some promising translation results. Moreover, the BLEU score results revealed that the Indonesian language is the best pivot language over any other examined Asian languages.**

Index Terms—**Pivot machine translation, Phrase-based statistical machine translation, Online machine translation services, English-Burmese machine translation, Low resources language pairs**

I. INTRODUCTION

THE statistical machine translation (SMT) is used for machine learning task of designing and developing statistical models and algorithms to translate texts from a source language into a target language. Parallel corpora are collection of text in one language and their equivalent translation to other languages. In machine translation research area, some language pairs include a huge number of parallel corpus which are easy to receive and ready to use. On the contrary, a low resource language pair is only a few of parallel corpus in small size or even not found at all. The rare of parallel corpus will directly cause to a bad translation. Burmese is a low resource language and the parallel corpus for Burmese to other languages such as Asian languages are not easy to find.

Moreover, pivot machine translation is to construct machine translation systems for a language pair where the availability of its parallel corpus (X-Z) is either absent or comparably smaller than the parallel corpora paired with a 'pivot' language 'Y', i.e., the X-Y and Y-Z corpora. When the parallel corpus of X-Z is small, taking advantage of X-Y and Y-Z corpora is the main approach of translating sentences from X to Z. It is one of the enabling techniques to build machine translation system for low-resource languages. Several pivot approaches are sentence translation or transfer method, triangulation method and synthetic method.

Wint Theingi Zaw is with Language Understandig Lab, Myanmar and master candidate in Yangon Technological University, Corresponding author email: wint.wtgz@gmail.com.

Ye Kyaw Thu is with LST Lab., NECTEC, Thailand and IDRI, CADT, Cambodia, Corresponding author email: yktnlp@gmail.com.

In this paper, we focus on the situations where small parallel corpora is available. The data sparseness problem suffered by phrase-based statistical machine translation system (PBSMT) when trained on small parallel corpora. Therefore, the pivot machine translation (English - Asian languages – Burmese) for small corpora are analysed by utilizing free available machine translation services such as "Google Translate", "Yandex Translate", "Baidu Translate" and "Bing Translate". The experiments were conducted by using phrase-based SMT with parallel corpora to observe the performance of translation. Moreover, small parallel corpora are built from English to Asian languages and Asian languages to Burmese language pair based on five English-Burmese bilingual books such as Constitution of the Republic of the Union of Myanmar in 2008 [1] and report book for the circumstance of Myanmar after Nargit [3], etc.

II. RELATED WORKS

In this section, the related works from phrase-based SMT and pivot machine translation are presented.

Chao-Hong Liu et al. [6] described that the pivot machine translation used the Chinese language as a pivot language. This paper presented the performance of pivot machine translation with the United Nations Parallel Corpus v1.0 (UN6Way) [7] based on pivot language such as English and Chinese. In addition, the experiments are analyzed for direct translation using SMT and NMT. Direct MT systems were still better than pivot MT system using the subset of 5,000K sentences pairs in UN ways corpus.

Zar Zar Hlaing et al. [8] presented that the SMT and NMT performance analyzed the increasing corpus extension using Google, SYSTRAN and Yandex Translate in online machine translation services. Firstly, the existing

corpus was extended using three freely available online machine translation services for English (en) and Thai (th) language pair. The performance of three statistical machine translation and neural machine translation was compared based on the original ASEAN-MT corpus and the extended version which is the double size of the original corpus. The experiment results showed that the performance of extended corpus improved 2.6% for th-en, 4.2% for en-th using SMT and 5.5% using NMT.

Sari Dewi Budiwati et al. [9] applied that statistical machine translation used multiple pivots for low resource languages. The experiments investigated many combinations of four phrase tables of multiple pivots on low resource language pairs in phrase-based SMT. The pivot languages were used as English, Myanmar, Malay and Filipino. The results showed the combination of multiple pivots outperformed the baseline.

Michael Paul et al. [10] presented that pivot language selection was important case for statistical machine translation. The aim of this study was to support next research on machine translation between under-resourced language pairs. Therefore, the suitable languages other than English used as pivot language. Pivot translation experiments using SMT techniques are carried out to translate between twelve languages which improved the translation quality of 61 language pairs when an English language was not used as the pivot language.

III. Parallel Corpora Building

Burmese (Myanmar language) is a low resource language, as the first step, the small parallel corpora of five books from English to Burmese were built. These corpora have 1,998 lines in Constitution of the Republic of the Union of Myanmar in 2008 [1], 431 lines in Gender and Political Participation in Myanmar [2], 3,481 lines in Meandering to Recovery (Post-Nargis Social Impacts Monitoring Ten Years After) [3], 1,133 lines in Parliamentary Research Training Manual [4] and 2,136 lines in Performance Analysis for State and Region Hluttaws of Myanmar [5].

IV. Methodology

In this section, the methodology of phrase-based SMT and the free online machine translation services used for the process of pivot translation from English to Burmese are described.

A. Phrase-based Statistical Machine Translation (PB-SMT)

Phrase based translation system works on the phrasal units [11]. A phrase is a simple sequence of words and not a linguistically motivated phrase. A phrase-based translation model typically achieves better performance than word-based models. A simple phrase-based translation model consists of phrase-pair probabilities that are extracted from the corpus and a basic reordering model,

and algorithm to extract the phrases to build a phrase-table [12].

The phrase translation model is built on noisy channel model. The best translation e_{best} that maximizes the translation probability $P(e|f)$ given the source sentences. Here, French (f) is the source language and an English (e) is the target language. The translation of a source language as French sentence (f) into target language as English sentence (e) is modeled as:

$$e_{best} = argmax_e P(e|f) \tag{1}$$

Applying the Bayes' rule, we can factorized the $P(e|f)$ into three parts.

$$P(e|f) = \frac{P(e)}{P(f)} P(f|e) \tag{2}$$

The final formulation of phrase-based model is as follows:

$$argmax_e P(e|f) = argmax_e P(f|e) P(e) \tag{3}$$

The standard phrase-based model has two components, the translation model $P(f|e)$ and the language model $P(e)$. However, the translation model actually can be split into two models, the phrase-translation model $\phi(\overline{f}_i, \overline{e}_i)$ and the distortion or reordering model $d(start_i - end_{i-1} - 1)$.

B. Pivot Machine Translation Using Free Online Machine Translation services

Pivot Machine translation uses to build English to Burmese systems without parallel data of the English-Burmese language pair. Asian languages as 'pivot' languages were used to build English to Burmese machine translation by using free online machine translation services. "Google Translate", "Yandex Translate", "Baidu Translate" and "Bing Translate" were used to translate the source to pivot languages and pivot languages to target language as in Fig.1. In this case, pivot languages owned from the translation of English using free online machine translation services.

1) *Google Translate*, a free machine translation service was developed by Google. This service can support to translate text and websites from one language to another languages. Google Translate launched in April 2006 with a statistical machine translation service, it used English as a pivot language instead of source to target direct translation. In November 2016, it changed to a neural machine translation engine named Google Neural Machine Translation (GNMT) which translates "whole sentences" at a time [13]. Google translate is used over 500 million users [14]. However, only 5,000 maximum number of characters are allowed per translation.

2) *Yandex Translate* was implemented by Yandex which is a free online machine translation web service [15]. It was lunched as beta mode of translation

in 2011. The service can support 96 languages and 10,000 maximum numbers of characters per translation. According to our knowledge, a hybrid machine translation system which combines neural and statistical approaches was utilized by Yandex Translate to receive high quality machine translation [16]. Yandex translation is more accurate than the other online machine translation services especially with Russian language, either as source or target language.

3) *Baidu Translate* is an online translation service launched in February 2011 by Baidu [17]. It supports nearly 200 languages. Users from different countries of the world easily access the information in Chinese and get in touch with the Chinese local service providers. Baidu Translate is based on neural machine translation and add a completely different method from MOSES, the most popular open source phrase-based machine translation system to obtain the better results.

4) *Bing Translate* (previously known as Windows Live translator) is a free online machine translation service provided by Microsoft. It was launched in 2007 and provides free text and website translations on the web. This service supports 73 language systems and 79 different language systems in October 2020. It used phrase-based SMT, syntax-based SMT and updated with neural machine translation as a default method of translating in May 2018. It allows 5,000 maximum numbers of characters per translation.

V. EXPERIMENTS

A. Corpora Statistics

The experiments were conducted based on two types of corpora. The first one is constructed with five small parallel corpora (2008 Constitution, Gender and Political, Nargit, Parliament research and Performance analysis) as mentioned in section III and the second is the combination of these five parallel corpora as the one corpus that includes 9,179 lines. In these experiments, the corpora are divided into 80% of the corpora for training data set, 10% of the corpora for development data set and remaining 10% for test data set.

B. Syllable Segmentation

In these experiments, the translated outputs from free online machine translation services are not available in exact and meaningful Burmese sentences and it is impossible to use word level segmentation for these outputs. Therefore, the translated outputs of these services are converted into syllable unit and PBSMT is trained using syllable unit segmentation for Burmese (Myanmar).

Burmese words are basically constructed the combination of multiple syllables, and most of the syllables are composed of more than one character. The structure of the syllable can be presented with Backus normal form (BNF) as following:

$$Syllable := CMV[CK][D] \qquad (4)$$

Here, consonants mean C, medial is M, vowel is V, vowel killer character is K, and diacritic characters is D. A rule-based approach, finite state automation (FSA) or regular expressions (RE) are used to segment Myanmar syllable (https://github.com/ye-kyaw-thu/sylbreak).

C. Moses SMT

PBSMT system provided by the Moses toolkit which are used for training in these experiments [18]. GIZA++ [19] was used to align the word segmented of source language with target language. The alignment was symmetrized by grow-diag-final and heuristic. The lexicalized reordering model was trained with the msd-bidirectional-fe option [20]. KenLM [21] was used in these experiments for training the 5-gram language model with modified Kneser-Ney discounting [22]. Minimum error rate training (MERT) [23] was used for tuning the decoder parameters and the decoding was done by using the Moses decoder (version 2.1.1) [18]. The experiments used default settings of Moses for all experiments.

D. Sentence Translation or Transfer Method in Pivot Machine Translation

The transfer method first translates the source language into pivot language using source pivot translation system, and then from pivot language to target language through the pivot target translation system. In this section, the process of pivot machine translation and pivot languages are described.

1) *Yandex Translate*: the English sentences are translated into pivot languages using Yandex Translate and the output sentences of the pivot languages are translated into Burmese using Google Translate as in Fig.1. Ten Asian languages such as Chinese, Hindi, Indonesian, Japanese, Khmer, Lao, Malay, Mongolian, Sinhala, and Thai were selected as the pivot languages in Yandex Translate.

2) *Baidu Translate*: the English sentences are translated into pivot languages using Baidu Translate and the output sentences of the pivot languages are translated into Burmese using Google Translate as in Fig.1. Nine Asian languages such as Chinese, Hindi, Indonesian, Japanese, Khmer, Lao, Malay, Sinhala and Thai were chosen as the pivot languages in Baidu Translate. In this case, Baidu Translate cannot support for Mongolian.

3) *Bing Translate*: the English sentences are translated into pivot languages using Bing Translate and the output sentences of the pivot languages are translated into Burmese using Google Translate

as in Fig.1. Six Asian languages such as Chinese, Hindi, Indonesian, Japanese, Malay and Thai were chosen as the pivot languages in Bing Translate. In this case, Baidu Translate cannot support for Khmer, Lao, Mongolian and Sinhala.

Fig. 1: The pivot machine translation using free online machine translation services

Burmese cannot be supported in every free online machine translation service especially in Baidu and Bing Translate in this study. Google and Yandex Translate can translate different language to Burmese in both directions. Although Google Translate cannot give 100% accuracy, it can better translate for Burmese than the Yandex Translate according to the knowledge of the Burmese's native speakers. Therefore, the experiments were designed to translate from pivot language into Burmese using Google Translate.

E. Evaluation

BLEU, RIBES and chrF++ were used to evaluate the machine translation outputs in these experiments. These are the de facto standard automatic evaluation metric Bilingual Evaluation Understudy (BLEU) [24], the Rank-based Intuitive Bilingual Evaluation Measure (RIBES) [25] and character n-gram F score (chrF++) [26]. The BLEU score determines the precision of n-gram (overall $n \leq 4$ in our case) with respect to a reference translation with a penalty for short translations. Intuitively, the BLEU score measures the adequacy of the translation and largest BLEU scores are best. RIBES is an automatic evaluation metric which based on rank correlation coefficients modified with precision and special care is paid to word order of the translation results. The RIBES score is appropriate for distance language pairs such as English and Burmese. The larger the RIBES scores, the better the translation quality. chrF++ [26] is an automatic evaluation which is based on character n-gram precision and recall enhanced with word n-grams. The F-score calculates average on all character and word, where the default order of word n-gram is 2 and character n-gram is 6.

VI. RESULT AND DISCUSSION

The BLEU score of PBSMT and pivot machine translation using free online machine translation services based on five parallel corpora are shown in Figures 2,3,4 and 5. Figure 2 shows the BLEU score of PBSMT in both directions. According to the result in Figure 2, the "2008 Constitution" corpus achieved the highest BLEU score (47.31 in English-Burmese and 31.12 in Burmese-English)

among other corpora in both directions. Interestingly, the results with each parallel corpus indicate that English to Burmese translation is better performance than Burmese to English translation direction.

The BLEU scores of pivot machine translation using free online machine translation services based on five parallel corpora are shown in Figure 3 for Yandex Translate, Figure 4 for Baidu Translate and Figure 5 for Bing Translate. Comparison of pivot languages, Indonesian and Hindi in Yandex Translate, Chinese and Indonesian in Baidu Translate, and Indonesian and Malay in Bing Translate described the better score in each parallel corpus. Moreover, Bing Translate in these free online machine translation services got the highest BLEU scores such as 37.00 in "2008 Constitution" corpus, 28.38 in "Gender and Political" corpus, 22.56 in "Nargit" corpus, and 19.41 in "Performance Analysis" corpus except 31.38 BLEU score in "Parliamentary Research" corpus with Yandex Translate. On the other hand, Bing Translate cannot support all selected 10 Asian languages as pivot languages like other translation services in these experiments. It can provide six languages (Chinese, Hindi, Indonesian, Japanese, Malay, Thai) in the selected pivot languages. In Figure 4, pivot languages as Khmer and Lao achieved lower scores than other languages in Baidu. In these experiments, "2008 Constitution" and "Parliamentary Research" achieved the best scores among five parallel corpora.

We have selected that the highest BLEU score of each parallel corpora conducted with the pivot machine translation from Figure 3,4 and 5 to compare the PBSMT result. Table. I shows that the BLEU score of PBSMT and pivot machine translation based on five parallel corpora. PBSMT received better BLEU score in 2008 constitution and Performance Analysis. The pivot machine translation using Bing translate got higher performance than PBSMT on other three corpus. Interestingly, pivot machine translation provide better translation than PBSMT on the smallest corpus (431 lines in Gernder and Political corpus).

The next experiment is based on the combined corpus (9,179 lines). The BLEU scores of pivot machine translation using free online machine translation services are shown in Figure 6. In this figure, Indonesian in pivot languages achieved the highest BLEU score in each free online machine translation service such as 26.99 in Yandex Translate, 26.47 in Baidu Translate and 27.29 in Bing Translate. The Figure 6 also shows that Malay language provide the second highest score for two online machine translation services (Baidu and Bing); therefore, we can consider it as a good pivot language for English to Burmese translation. Table. II shows the BLEU, RIBES and chrF++ scores of PBSMT and the highest result of pivot machine translation in Figure 6. In Table. II, PBSMT achieved the highest BLEU and chrF++ scores than pivot machine translation but pivot machine translation got the best RIBES score. Of the three online machine translation services, Yandex Translate achieved the highest RIBES score, while Bing Translate achieved the highest BLEU score.

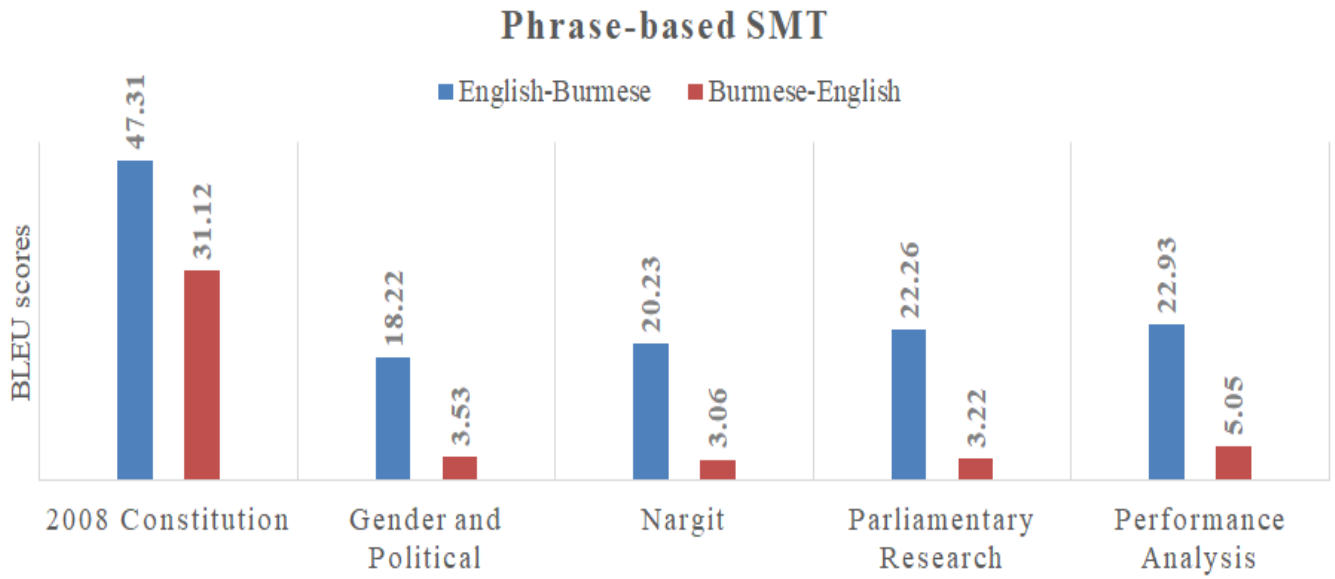Table. III shows some translation examples of PBSMT

Phrase-based SMT

Fig. 2: The BLEU scores of Phrase-based SMT based on the five parallel corpora in both directions
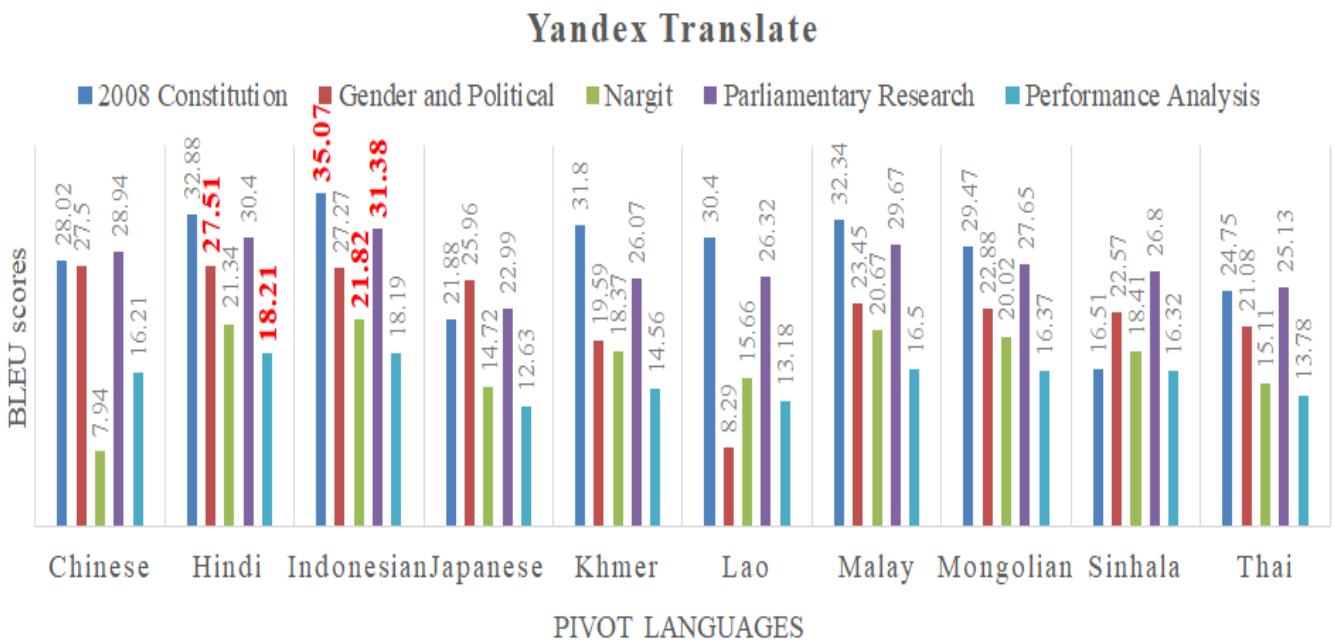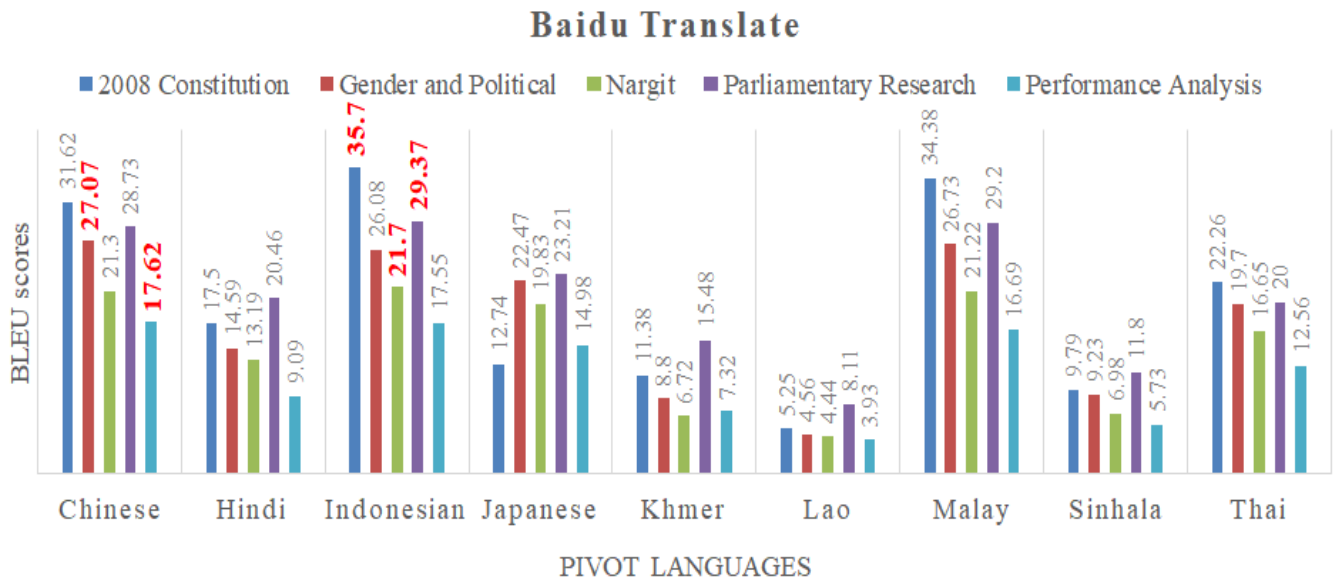
Yandex Translate

Fig. 3: The BLEU scores of English-to-Burmese pivot machine translation using "Yandex Translate" for each parallel corpus

TABLE I: The BLEU scores of PBSMT and Pivot Machine Translation for five construced parallel corpora

Corpus	PBSMT	Pivot Machine Translation		
		Yandex	Baidu	Bing
2008 Constitution	**47.31**	35.07	35.7	35.79
Gender and Political	18.22	**27.51**	**27.07**	**28.38**
Nargit	20.23	21.82	21.7	**22.56**
Parliamentary Research	22.26	31.38	29.37	**30.06**
Performance Analysis	**22.93**	18.19	17.62	19.41

Baidu Translate

Fig. 4: The BLEU scores of English-to-Burmese pivot machine translation using "Baidu Translate" for each parallel corpus
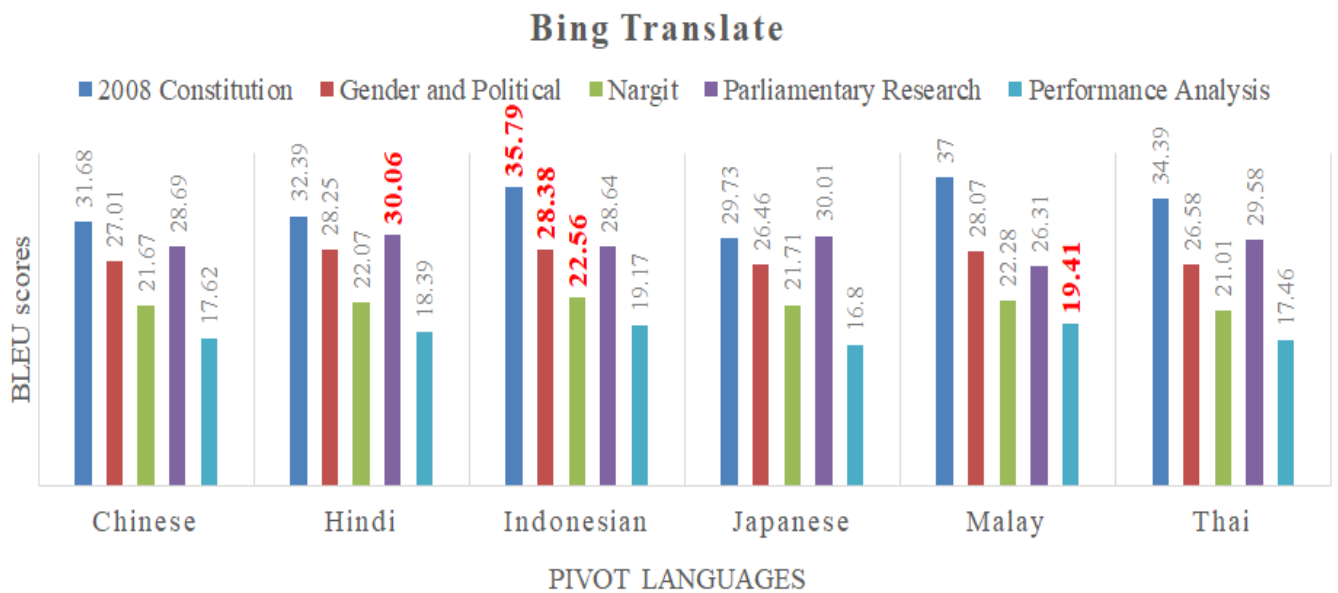
Bing Translate

Fig. 5: The BLEU scores of English-to-Burmese pivot machine translation using "Bing Translate" for each parallel corpus

TABLE II: The BLEU, RIBES and chrF++ scores of PBSMT and Pivot Machine Translation for the combined corpus

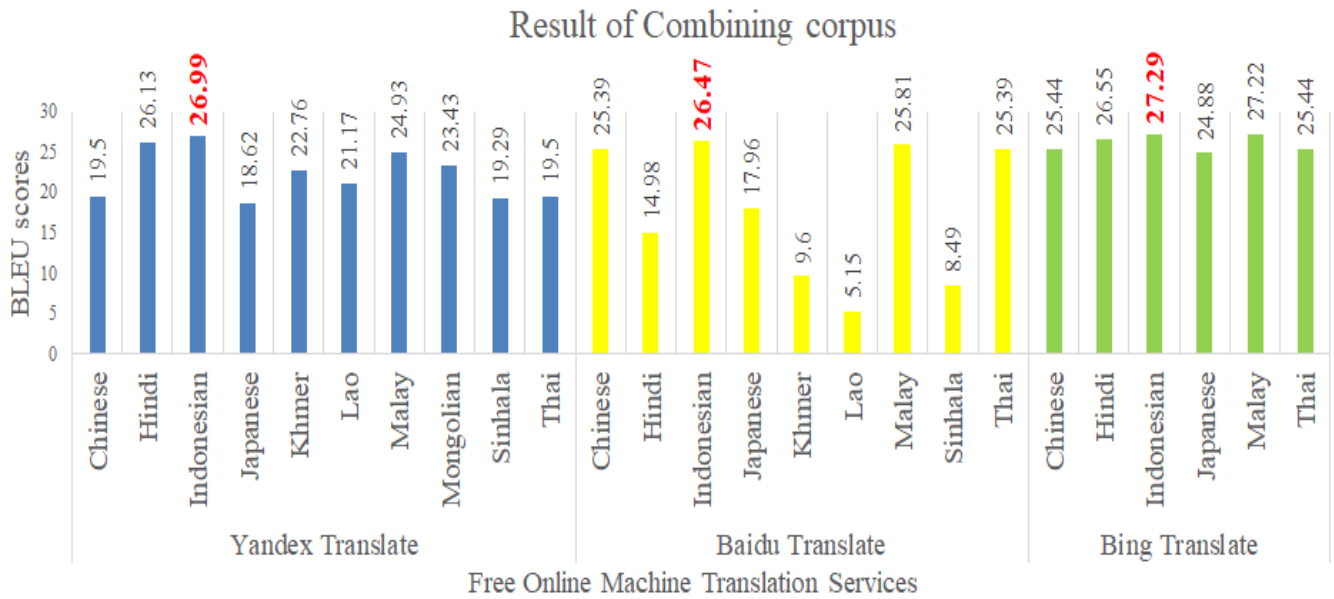| Combine Corpus	BLEU	RIBES	chrF++	
PBSMT (English-Burmese)	**28.33**	0.62977		**45.3354** **43.5437**
Yandex (English-Indonesian-Burmese)	26.99	**0.700007**	c6+w2-F2 c6+w2-avgF2	44.7449 43.4930
Baidu (English-Indonesian-Burmese)	26.47	0.692384		43.6916 42.1431
Bing (English-Indonesian-Burmese)	27.29	0.698874		44.8175 43.2448

Fig. 6: The BLEU scores of pivot machine translation with the combined parallel corpus

TABLE III: Examples of English-Burmese Translation with PBSMT and Pivot Machine Translation Using Indonesian as a Pivot Language

Machine Translation	Source (English)	1 7 2 . (a) The first regular session of the Region or State Hluttaw shall be convened by the State Peace and Development Council after the Constitution comes into operation.
	Reference (Burmese)	၁ ၇ ၂ ။ (က) ဖွဲ့ စည်း ပုံ အ ခြေ ခံ ၃ ၈ ဒေ အာ ဏာ တည် ပြီး နောက် ပ ထ မ အ ကြိမ် တိုင်း ဒေ သ ကြီး သို့ မ ဟုတ် ပြည် နယ် လွှတ် တော် ပုံ မှန် အ စည်း အ ဝေး ကို နိုင် ငံ တော် အေး ချမ်း သာ ယာ ရေး နှင့် ဖွံ့ ဖြိုး ရေး ကောင် စီ က ခေါ် ယူ ကျင်း ပ ရ မည် ။
PBSMT	Target: Burmese	၁ ၇ ၂ ။ (က) နိုင် ငံ တော် ၏ ပ ထ မ အ ကြိမ် ပုံ မှန် အ စည်း အ ဝေး များ ကို တိုင်း ဒေ သ ကြီး သို့ မ ဟုတ် ပြည် နယ် လွှတ် တော် အ စည်း အ ဝေး ခေါ် ယူ မှု အ နည်း အ များ ပြည် နယ် ဒီ မို့ က ရေ စီ အ ပြီး အ ခြေ ခံ ၃ ၈ ဒေ အ တည် ပြု ပြဋ္ဌာန်း ဆောင် ရွက် ရ မည် ။
Pivot Machine Translation: (Source → Yandex → Pivot → Google → Target)	Pivot: Indonesian	1 7 2 . (a) pertama sesi reguler dari Wilayah atau Negara Hluttaw harus diselenggarakan oleh the State Peace and Development Council setelah Konstitusi datang ke dalam operasi .
	Target: Burmese	1 7 2 ။ (က) လွှတ် တော် နယ် နိ မိတ် သို့ မ ဟုတ် ပြည် နယ် ၏ ပ ထ မ အ ကြိမ် ပုံ မှန် အ စည်း အ ဝေး ကို ဖွဲ့ စည်း ပုံ အ ခြေ ခံ ၃ ၈ ဒေ အ ကောင် အ ထည် ဖော် ပြီး နောက် နိုင် ငံ တော် အေး ချမ်း သာ ယာ ရေး နှင့် ဖွံ့ ဖြိုး ရေး ကောင် စီ က အိမ် ရှင် အ ဖြစ် လက် ခံ ကျင်း ပ ရ မည် ။
Pivot Machine Translation: (Source → Baidu → Pivot → Google → Target)	Pivot: Indonesian	1 7 2.a) Sesi biasa pertama dari wilayah atau negara Hluttaw akan dipanggil oleh Dewan Perdamaian dan Pembangunan Negeri setelah Konstitusi mulai bekerja .
	Target: Burmese	၁ ၇ ၂။ က) ဖွဲ့ စည်း ပုံ အ ခြေ ခံ ၃ ၈ ဒေ စ တင် ပြီး ပါ က တိုင်း ဒေ သ ကြီး သို့ မ ဟုတ် ပြည် နယ် ၏ ပ ထ မ အ ကြိမ် ပုံ မှန် အ စည်း အ ဝေး ကို နိုင် ငံ အေး ချမ်း သာ ယာ ရေး နှင့် ဖွံ့ ဖြိုး ရေး ကောင် စီ က ဆင့် ခေါ် လိမ့် မည် ။
Pivot Machine Translation: (Source → Bing → Pivot → Google → Target)	Pivot: Indonesian	1 7 2 . (a) Sesi reguler pertama Wilayah atau Negara Hluttaw akan dis-elenggarakan oleh Dewan Perdamaian dan Pembangunan Negara setelah Konstitusi mulai beroperasi.
	Target: Burmese	1 7 2 ။ (က) လွှတ် တော် နယ် နိ မိတ် သို့ မ ဟုတ် ပြည် နယ် ၏ ပ ထ မ အ ကြိမ် ပုံ မှန် အ စည်း အ ဝေး ကို ဖွဲ့ စည်း ပုံ အ ခြေ ခံ ၃ ၈ ဒေ ပေါ် ပေါက် လာ သည် နှင့် တ ပြိုင် နက် နိုင် ငံ တော် အေး ချမ်း သာ ယာ ရေး နှင့် ဖွံ့ ဖြိုး ရေး ကောင် စီ က အိမ် ရှင် အ ဖြစ် လက် ခံ ကျင်း ပ မည် ဖြစ် သည်။

and pivot machine translation using Yandex, Baidu and Bing Translate based on the combined corpus. Indonesian is presented as one example of translated pivot languages in this table because the Indonesian achieved the highest score on the experiments of combined corpus. Comparison of the translated Burmese sentence using free online machine translation services in Table. III, Yandex and Bing additionally translate the extra words which is the unnecessary words as "အိမ် ရှင် အ ဖြစ်" (as a host country). In addition, Yandex and Bing cannot correctly translate Burmese number, but Baidu provide it such as "1 7 2." (English) into "၁ ၇ ၂။" (Burmese).

VII. CONCLUSION

In this paper, we consider the real-world scenario of English-to-Burmese sentence level pivot translation without a parallel corpus. The combinations of freely available machine translation services Yandex-Google, Baidu-Google, and Bing-Google were done with Asian pivot languages. The small five English-Burmese parallel corpora were developed for evaluation. Bing Translate shows better results than Yandex and Baidu translation services in this experiment. Generally, for Phrase-based SMT with the parallel corpus achieved better BLEU and chrF++ scores, some pivot machine translations without parallel corpus achieved better RIBES scores and acceptable BLEU and chrF++ scores. Pivot machine translation achieved better performance on the smallest corpus. However, there remains a major challenge the error propagation problem of the pivot approach: the errors made in the source-to-pivot translation will be propagated to the pivot-to-target translation.

REFERENCES

[1] https://www.burmalibrary.org/sites/burmalibrary.org/files/obl/docs09/Myanmar_Constitution-2008%28en%26bu%29-red.pdf

[2] https://emref.org/sites/emref.org/files/publication-docs/gender_and_political_in_myanmarenglish_online.pdf

[3] https://emref.org/sites/emref.org/files/publication-docs/post-nargis_social_impact_monitoring_10_year_aftersim_5_eng_version.pdf

[4] https://emref.org/sites/emref.org/files/publication-docs/parliamentary_research_training_manual_eng.pdf

[5] https://emref.org/sites/emref.org/files/publication-docs/performance_analysiseng_2nd_edition_final.pdf

[6] Liu, Chao-Hong, Catarina Cruz Silva, Longyue Wang, and Andy Way. "Pivot machine translation using Chinese as pivot language." In China Workshop on Machine Translation, Springer, Singapore, 2018, pp. 74-85.

[7] Ziemski, Michał, Marcin Junczys-Dowmunt, and Bruno Pouliquen. "The united nations parallel corpus v1. 0." In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2016, pp. 3530-3534.

[8] Zar Zar Hlaing, Ye Kyaw Thu, Thepchai Supnithi, Ponrudee Netisopakul, "Increasing SMT and NMT Performance by Corpus Extension with Free Online Machine Translation Services", In Proceedings of the 4th International Conference on Advanced Information Technologies (ICAIT-2020), Nov 4-5 2020, Yangon, Myanmar, pp. 212-217.

[9] Budiwati, Sari Dewi, and Masayoshi Aritsugi. "Multiple Pivots in Statistical Machine Translation for Low Resource Languages." Proceedings of the 33rd Pacific Asia Conference on Language, Information and Computation. Waseda Institute for the Study of Language and Information, 2019.

[10] Paul, Michael, Hirofumi Yamamoto, Eiichiro Sumita, and Satoshi Nakamura. "On the importance of pivot language selection for statistical machine translation." In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers, 2009, pp. 221-224.

[11] P.Koehn, F. J. Och, and D. Marcu, "Statistical phrase-based translation." in Proceedings of HTL-NAACL, 2003, pp. 48-54.

[12] Lucia Specia, "Tutorial, Fundamental and New Approaches to Statistical Machine Translation", International Conference Tecent Advances in Natural Language Processing, 2011.

[13] Google Tran 2006, "Google Translate", May 10, 2020, https://en.eikipedia.org/wiki/Google−Translate.

[14] Google Tran 2016, Turovsky, Barak (April 28, 2016) "Ten years of Google Translate", Google Translate Blog, https://en.eikipedia.org/wiki/ Google−Translate

[15] Yandex 2011, "Yandex Translate", May 10, 2020, https://yandex.com/company/technologies/translation/

[16] Yandex 2017, "Oe model is better than two. Yandex Translate launches a hybrid machine translation system", May 10, 2020

[17] He, Zhongjun. "Baidu translate: Research and products." In Proceedings of the Fourth Workshop on Hybrid Approaches to Translation (HyTra), 2015, pp. 61-62.

[18] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst, "Moses: Open Source Toolkit for Statistical Machine Translation", Annual Meeting of the Association for Computational Linguistics (ACL), demonstration session, Prague, Czech Republic, June 2007, pp.177-180.

[19] Och Franz Josef and Ney Hermann, "Improved Statistical Alignment Models", in Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, Hong Kong, China, 2000, pp.440-447.

[20] Tillmann Christoph, "A Unigram Orientation Model for Statistical Machine Translation", in Proceedings of HLT-NAACL 2004: Short Papers, Stroudsburg, PA, USA, 2004, pp. 101-104.

[21] Heafield, Kenneth, "KenLM: Faster and Smaller Language Model Queries", In Proceedings of the Sixth Workshop on Statistical Machine Translation, WMT 11, Edinburgh, Scotland, 2011, pp. 187-197.

[22] Chen Stanley F and Goodman Joshua, "An empirical study of smoothing techniques for language modeling", in Proceedings of the 34th annual meeting on Association for Computational Linguistics, 1996, pp. 310-318.

[23] Och Franz J., "Minimum error rate training in statistical machine traslation", in Proceedings of the 41 st Annual Meeting Association for Computational Linguistics – Volume 1, Association for Computer Linguistics, Sapporo, Japan, July 2003, pp.160-167.

[24] Papineni K, Roukos S, Ward T, Zhu WJ. "BLEU: a method for automatic evaluation of machine translation". In Proceedings of the 40th annual meeting of the Association for Computational Linguistics, July 2002, pp. 311-318.

[25] Isozaki, H., Hirao, T., Duh, K., Sudoh, K. and Tsukada, H. "Automatic evaluation of translation quality for distant language pairs". In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, October 2010, pp. 944-952.

[26] Popović, M. "chrF++: words helping character n-grams". In Proceedings of the second conference on machine translation, September 2017, pp. 612-618.

Wint Theingi Zaw received the B.E. degree in Information Technology at Technological University (Pathein) in 2008. She is also a Master candidate in Computer Engineering and Information Technology (CEIT) Department of Yangon Technological University (YTU). She focused her efforts on her Master thesis research of Machine Translation Based Myanmar Romanization.

Ye Kyaw Thu is a Visiting Professor of Language & Semantic Technology Research Team (LST), Artificial Intelligence Research Unit (AINRU), National Electronic & Computer Technology Center (NECTEC), Thailand and Affiliate Professor at Cambodia Academy of Digital Technology (CADT), Cambodia. He is also a founder of Language Understanding Lab., Myanmar. His research lie in the fields of artificial intelligence (AI), natural language processing (NLP) and human-computer interaction (HCI). He is actively co-supervising/supervising undergrad, masters' and doctoral students of several universities including Assumption University (AU), Kasetsart University (KU), King Mongkut's Institute of Technology Ladkrabang (KMITL) and Sirindhorn International Institute of Technology (SIIT).

Zar Zar Hlaing received a bachelor's degree in computer science from the University of Computer Studies, Monywa, in 2014 and a master's degree from the University of Computer Studies, Mandalay, in 2019. She is currently a Ph.D. research student in the School of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. She is also doing research in Machine Translation focusing on low-resource languages such as Thai and Myanmar. Her research interests include natural language processing (NLP), software engineering, and big data analytics.

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