Statistical Machine Translation between Myanmar (Burmese) and Chin (Mizo) Language

Hnin Yi Aye
Department of CEIT
Yangon Technological University
Yangon, Myanmar
hyaye13itphd@ytu.edu.mm

Yuzana Win
Department of CEIT
Yangon Technological University
Yangon, Myanmar
yuzanawinn@ytu.edu.mm

Ye Kyaw Thu
LU Lab., Myanmar
LST, NECTEC, Thailand
yktnlp@gmail.com

***Abstract*—The main motivation of this paper is to develop the statistical machine translation between Myanmar (Burmese) and Chin (Mizo) language. In this paper, there are three main parts: parallel corpus creation, applying statistical machine translation approaches and evaluation for error analysis. *Firstly*, we develop a parallel corpus of around 15K sentences based on the ASEAN-MT dataset. *Secondly*, we apply three statistical approaches: phrase-based (PBSMT), hierarchical pharse-based (HPBSMT) and operation sequence model (OSM). *Finally*, we use the evaluation metrics (BLEU, RIBES, chrF scores) and word error rate (WER) analysis to measure the translation performance. The experimental results show that OSM approach can obtain not only the highest BLEU scores in both directions but also RIBES and chrF^{++} scores in Chin-Myanmar direction. Furthermore, we investigate the error analysis and confusion pairs of machine translation among three approaches.**

***Index Terms*—Myanmar-Chin(Mizo)Machine Translation, PBSMT, HPBSMT, OSM, WER, BLEU**

I. Introduction

Statistical machine translation (SMT) systems are required large amount of data to achieve the good performance in translation process. Although the state-of-the-art techniques of statistical machine translation [2] were evaluated for many rich resource languages such as English, French or German, there is still a few translation system for low-resource language pair [3]. This is because it is difficult to acquire a large amount of parallel corpora for under-resourced languages like Myanmar.

In recent years, there have been some studies on the machine translation of Myanmar language. Ye Kyaw Thu et al. [4] contributed the first large scale evaluation of the performance of three SMT approaches (PBSMT, HPBSMT and OSM) between Myanamar and twenty other languages, in both directions. In their experiment, syllable and supervised word segmentation schemes were applied. As a result, the highest quality machine translation was attained with supervised CRF-based word segmentation in all of the experiments and the HPBSMT and OSM approaches were achieved the highest BLEU and RIBES scores.

Win Pa Pa et al. [5] also proposed a statistical machine translation methods for English, Thai, Lao and Myanmar languages. They used PBSMT, HPBSMT, tree-to-string (T2S), string-to-tree (S2T) and OSM methods for translation. Their experimental results showed that HPBSMT approach was achieved the highest BLEU and PBSMT is attained the best RIBES scores. Ye Kyaw Thu et al. [6] presented a large-scale study of statistical machine translation methods (PBSMT, HPBSMT and OSM) for Khmer language. In their experiment, syllable and supervised word segmentation schemes were applied. As a result, HPBSMT and OSM were obtained the highest level in both BLEU [25] and RIBES scores for distant languages and OSM for other non-distant languages.

From the above analysis, there is no publicly available parallel corpus for under-resourced languages, like Myanmar and Chin. The Chin (Mizo) language is one of the main ethnic languages in Myanmar and it is often considered as under-resourced language. Therefore, the main contribution of this paper is to develop a newly language pair Myanmar-Chin parallel corpus. Moreover,the three statistical machine translation approaches such as PBSMT, HPBSMT and OSM are used to measure the performance of translation in both directions. This is because, above three approaches could achieve the good translation for non-distant languages such as Myanmar to English, Khmer and twenty others languages in both directions.

The rest of this paper is organized into seven sections. Section 2 describes the related work with the translation system. Section 3 introduces the Chin language. Section 4 explains the methodology of the applied approaches. Section 5 includes the experiments for Myanmar-Chin translation process. Section 6 performs the evaluation measure. Section 7 presents the word error analysis and the final section concludes the discussion on future work.

II. Related Work

In this section, the state-of-the-art relating to the translation systems for Myanmar and other languages are

978-1-7281-9896-5/20/$31.00 ©2020 IEEE

described.

Thazin Myint Oo et al. [7] presented the machine translation between Myanmar (Burmese) and Rakhine (Arakanese) languages. They used PBSMT as the baseline system, and the other statistical methods HPBSMT and OSM were also investigated. Both BLEU and RIBES scores for all approaches were comparable and high-ranked for both Rakhine to Myanmar and Myanmar to Rakhine translations. According to the experimental results, OSM approach could obtain the highest of BLEU and RIBES scores.

Thet Thet Zin et al. [8] also proposed Myanmar phrases translation model with morphological analysis based on statistical approach. According to their results, morphologically analysis on Myanmar language could be improved the translation quality.

Amarnath Pathak et al. [11] described a method for English to Mizo translation by using both neural and statistical (PBSMT) approaches. According to their experimental results, the proposed approaches could achieved fair and comparable BLEU score about 22.

III. Chin (Mizo) Language

Chin is one of the main seven ethnic minority states located in the Republic of the Union of Myanmar. Chin State previously known as Chin Hills was split from Arakan State in 1948 and officially named as Chin State in 1974. The Chin state is located in western Myanmar and it is a highland. The area of Chin state is 36,018.8 km2 (13,906.9 sq mi) and difficult for transportation and underpopulated. It still remains as one of the least developed areas in Myanmar.

There are a total of eight ethnic groups in Chin State. Laimi, Matu, Zomi, Mizo, Asho, Mara, Khumi and Daai are the popular main languages that belong to Chin ethnic groups. The mizo, Mizo Twang which is also known as Lusei or Lushai is spoken as a native language by Mizo people not only in the Chin state of Myanmar but also in Mizoram State of India and Bangladesh. It can be regarded as the Kuki-Chin language belonging to the Sino-Tibetan language family. The total population of Mizo people is 200,000 in Myanmar, 1,200,000 in India, and 1,500 in Bangladesh.

Like Myanmar language, the Mizo language is also a tonal language. It has eight tones for each vowel especially four for reduced tones and four for long tones. Therefore, Mizo and Myanmar (Burmese) languages have the same pronunciation in some few words. For example, kam (“ကမ်း”) (“bank of a river”), kun (“ကွေး”) (“bend”), kha (“ခါ”) (“bitter”), mei (“မီး”) (“fire”), ni (“နေ”) (“sun”), hnih (“နှစ်”) (“two”), li (“လေး”) (“four”), nga (“ငါး”) (“five”), sam (“ဆံ”) (“hair”), that (“သတ်”) (“kill”) [13].

The Mizo alphabet is adopted from the Roman script and it has 25 main letters. They are similar to English alphabets and some are combination of two English letters i.e “Aw”, “Ch” and “Ng”. The pronunciation of Mizo alphabet is also the nearly same with English except for the Mizo letters “a, aw, ch, e, and T”. For example, the Mizo alphabet “Aw” is pronounced as in English alphabet “O” (/ɔ/oː/), “E” is nearly pronounced as in English letter “A” (/eɪ/), and also “A” as “R” (/ɑː/ɑr/) in English. The declarative word order is Object-subject-verb (OSV) structure.

In contrast to Myanmar language, it is exactly different in script style, word order and segmentation rules. Some examples for parallel sentences of Myanmar (my) and Mizo Chin (ch) are described as follows:

1) ကျွန်တော် နောက် အစည်းအဝေး တစ်ခုကို တက် မှာ သေချာပါတယ် ။
   Meeting leh-pek pawh ka tel ngei dawn e .
   I am sure to attend the next meeting.

2) မင်း မိဘ တွေ ကို ကြည့်ရှုစောင့်ရှောက် သင့်တယ် ။
   I nu leh pa i enkawl a .
   You should take care of your parents.

In Mizo language, some suffixes are used to distinguish the gender. For example, in animals, suffix “-nu” means female, and “-pa” means male. For human, suffix “-a” in name refers to male and “-i” refers to female.

IV. Methodology

In this section, the famous statistical machine translation methods called PBSMT, HPBSMT and OSM are applied. The baseline system is the phrase-based statistical machine translation [1].

A. Phrase-Based Statistical Machine Translation (PBSMT)

The main components of phrase-based translation are translation model, language model and decoding model. The translation model is based on phrasal units. Here, a phrase is a simple, contiguous sequence of words and there is no need to be linguistically motivated. Reordering of translated target phrases is handled by a distance-based reordering model. In machine translation, decoding is to find the best scoring translation and it is a challenging task since there is an exponential number of possible target hypotheses, given a specific input source sentence. A language model is an essential component of the decoding process, which measures the fluency and reasonable degree of a translated sentence structure.

The mathematical formulation for phrase-based translation model can be denoted as Eq. 1:

$$\hat{e} = arg \max_{e} P(e|f) \tag{1}$$

Applying the Bayes' Rule, we can factorized the $P(e|f)$ into three parts $P(e)$, $P(f)$ and $P(f|e)$ as shown in Eq. 2:

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

The final mathematical formulation of phrase-based model is in Eq. 3:

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

Fig. 1 illustrates the sample translation and reordering steps of the PBSMT for Myanmar to Chin (Mizo) language. The Myanmar input sentence is first segmented into so-called phrases (any multiword units). Then, each Myanmar phrase is translated into Chin phrase. Finally, the translated Chin phrases are reordered. In Fig. 1, the five Myanmar words and Chin words are mapped as five phrase pairs.

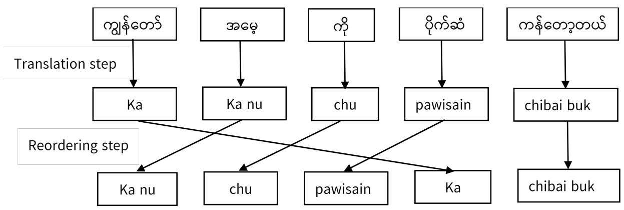

Fig. 1. An example translation and reordering steps of PBSMT for Myanmar to Chin (Mizo) language

B. Hierarchical Phrase-Based Statistical Machine Translation (HPBSMT)

The hierarchical phrase-based SMT approach is a model based on synchronous context-free grammar [9] and learned from the bilingual corpus with no syntactic annotations. It is a mixture of nonterminal and terminal symbols on the right side of a sentence. It encapsulates nicely the type of reordering involved when translating. It is suitable for long distance reordering in translation [10]. The grammatical example for hierarchical phrase-based translation between Myanmar and Chin language is as follows and the Burmese phrase “ကူညီ” means “help” in English.

[X][X] ကူညီ [X] ||| [X][X] a pui [X] |||
[X][X] ကူညီ [X] ||| [X][X] kan pui [X] |||
[X][X] ကူညီ [X] ||| [X][X] pui [X] |||
[X][X] ကူညီ [X] ||| [X][X] puih [X] |||
[X][X] ကူညီ [X] [X] [X] ||| [X][X] pui [X] [X] [X] |||

C. Operation Sequence Model (OSM)

The Operation Sequence Model [14] is a new SMT model that captures benefits of PBSMT and N-gram based SMT frameworks. In OSM approach, the translation and reordering steps are integrated into a single generative process. This approach is enable to memorize phrase pairs. The advantages of OSM over PBSMT are capturing source and target information across phrasal boundaries and avoiding any spurious phrasal segmentation ambiguity. A better, long distance reordering mechanism is the unique property of OSM. There are five operations for translation and three operations for reordering process.

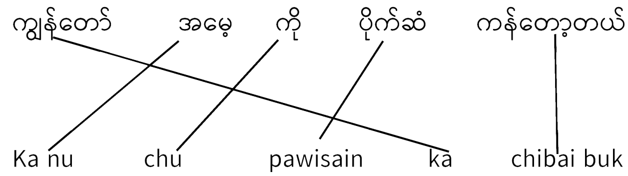

Fig. 2. An example of operation sequence translation from Myanmar to Chin (Mizo)

Fig. 2 describes the example translation of Myanmar sentence “ကျွန်တော် အမေ့ ကို ပိုက်ဆံ ကန်တော့တယ်” to Chin language by using OSM method. The detail operations are as follows:

Source: ကျွန်တော် အမေ့ ကို ပိုက်ဆံ ကန်တော့တယ်
Target: ka nu chu pawisain ka chibai buk
Operation 1: Insert Gap
Operation 2: Generate (အမေ့, Ka nu)
Operation 3: Generate (ကို, chu)
Operation 4: Generate (ပိုက်ဆံ, pawisain)
Operation 5: Jump Back (1)
Operation 6: Generate (ကျွန်တော်, ka)
Operation 7: Jump Forward
Operation 8: Generate (ကန်တော့တယ်, chibai buk)

The probability of a sequence of operations is defined according to an n-gram model, i.e., the probability of an operation depends on the $n-1$ preceding operations. Let $O = o_1, ..., o_n$ be a sequence of operations as hypothesized by the translator to generate a word-aligned bilingual sentence pair $< F; E; A >$; the model is then defined as:

$$P_{osm}(F, E, A) = P(O_1^J) = \prod_{j=1}^{J} p(o_j | o_{j-n+1}, ..., o_{j-1}) \tag{4}$$

V. Experiments

A. Corpus Statistics

We used around 15K Myanmar sentences (without name entity tags) of the ASEAN-MT Parallel Corpus. This corpus consists of the six main groups under the travel domain related to daily conversations, travelling purposes, shopping, cosmetics, communications, information for food, restaurant and hotel, transportation and health care system.

Translation into monolingual Chin corpus was manually done by native Chin people and the translated corpus was validated by the members of Mizo Chin association named “Yangon Zofa Thalai Pawl” (YZTP). Manual word segmentation for Chin language is also performed and there are totally 477,986 words in the corpus. For 10-fold cross-validation experiments, we prepared the training set of 14,976 to 14,998 sentences, development set of 500 to 507 sentences and 1,489 to 1,507 sentences for evaluation respectively.

B. Word Segmentation

Firstly, the parallel corpus is split into source and target files. Then, the corpus of both Myanmar and Chin languages are tokenized. A regular expression based Myanmar syllable segmentation tool named "sylbreak" is used to segment the sentences in Myanmar language of ASEAN-MT parallel corpus. The segmentation for Chin language is followed by the Bloomfieldian model and whitespace is used for word segmentation. Finally, the data corpus is cleaned by removing the extra space, empty lines that the lines have been too short or too long and the misaligned sentences of the two languages.

C. Moses SMT System

We applied the statistical models consists of the translation model, language model, and decoding model.These models are implemented by using the processing tools such as GIZA++ [15], KenLM, and Moses decoder [18]. GIZA++ was used for alignments of segmented words in the source and target languages. By a grow-diag-final and heuristic, the alignment was symmetrize and the msd-bidirectional-fe option [16], the lexicalized reordering model was trained. For training the n-gram language model with modified Kneser-Ney discounting [17], KenLM [20] was used. Minimum error rate training (MERT) [22] was applied for tuning the hyper-parameters and the decoding was done by using the Moses decoder (version 2.4.10) [18]. We applied Moses' default settings for all experiments, such as PBSMT, HPBSMT and OSM.

VI. Evaluation

The evaluation of experimental results is measured by four criteria, Bilingual Evaluation Understudy (BLEU) [24], Rank-based Intuitive Bilingual Evaluation Score (RIBES), character n-gram F-score (chrF^{++}) scores [28], [29] for accuracy and Word Error Rate (WER) [26] for the rate of error. BLEU is the de facto standard for evaluation.

BLEU score measures the n-gram precision with respect to the comparison of machine translated output (hypothesis) and the multiple reference translations with a brevity penalty for short sentences. As BLEU is the metric of accuracy, the opposite of error rate, the larger BLEU is the better performance. BLEU score is calculated by the following Eq. 5:

$$BLEU_n = (\text{brevity-penalty}) \prod\nolimits_{i=1}^{n} precision_i \qquad (5)$$

where brevity-penalty= $min(1, \frac{output\_length}{reference\_length})$

RIBES uses rank correlation coefficients based on word order to compare hypothesis and reference translations.

ChrF is calculated the F-score based on character n-grams. The general formula for chrF score is Eq. 6:

$$CHRF\beta = (1+\beta^2)\frac{CHRP.CHRR}{\beta^2.CHRP + CHRR} \qquad (6)$$

where CHRP and CHRR stand for character n-gram precision and recall arithmetically averaged over all n-grams. In our expeirment chrF^{++} scores are measured.

Word Error Rate (WER) [27] is computed as the minimum number of substitution, insertion and deletion operations that have to be performed to convert the hypothesis sentence into the reference sentence. WER can be computed by the following Eq. 7.

$$WER = \frac{(I+D+S)\times 100}{N} = \frac{(I+D+S)\times 100}{S+D+C} \qquad (7)$$

where S, I, D and C refers to number of substitutions, insertions, deletions, and correct words respectively. Here, the number of words in the reference sentences is denoted as N ($N = S + D + C$).

Tables I, II and III illustrate the average translation performances based on 10-fold cross-validation of PBSMT, HPBSMT and OSM measured in BLEU, RIBES and chrF scores respectively [23]. The best scores among three SMT approaches are described in bold. Here, "my" is the abbreviation for Myanmar, "ch" for Chin, "src" for source language and "trg" for target language respectively. According to tables I, II and III, OSM approach can obtain the highest BLEU scores in both directions as well as RIBES and chrF scores in Chin-to-Myanmar direction. Moreover, HPBSMT can also obtain the highest RIBES and chrF scores for Myanmar-to-Chin direction. Interestingly, the scores of all three methods are comparable performance. We found that our parallel corpus is more effective for Myanmar-to-Chin translation than Chin-to-Myanmar direction because of a better performance (around 10 in BLEU score).

Table IV shows the average percentage of WER values of PBSMT, HPBSMT and OSM methods for Myanmar-to-Chin and Chin-to-Myanmar translations with a test data of around 1.5K sentences (one-tenth of total training sentences).

TABLE I
Average BLEU scores for PBSMT, HPBSMT and OSM

src-trg	PBSMT	HPBSMT	OSM
my-ch	20.68	18.94	**21.16**
ch-my	11.19	11.67	**11.69**

TABLE II
Average RIBES scores for PBSMT, HPBSMT and OSM

src-trg	PBSMT	HPBSMT	OSM
my-ch	0.768	**0.770**	0.768
ch-my	0.646	0.650	**0.651**

To compare WER values of the three approaches, all are very closed values to each other. The lowest WER value is resulted from OSM and the highest is from PBSMT for both directions.

TABLE III
AVERAGE CHRF ++ SCORES FOR PBSMT, HPBSMT AND OSM

src-trg	chrF Score Type	PBSMT	HPBSMT	OSM
my-ch	c6+w2-avgF2	37.9199	**38.2767**	38.2390
	c6+w2-F2	35.3158	**35.7516**	35.7198
ch-my	c6+w2-avgF2	38.3276	37.8621	**38.8027**
	c6+w2-F2	**35.7694**	35.6578	36.7547

TABLE IV
AVERAGE PERCENTAGE OF WER FOR PBSMT, HPBSMT AND OSM (ABOUT 1500 SENTENCES)

src-trg	PBSMT	HPBSMT	OSM
my-ch	61.46%	61.45%	**61.44%**
ch-my	74.12%	73.90%	**73.78 %**

VII. Error Analysis

The NIST scoring toolkit (Version: 2.4.11) named SCTK is used for score calculation [21]. The SCLITE, a core program of SCTK, is used to align erroneous hypothesis sentences with error-free references sentences and calculate the word error rate (WER). Then SCLITE program is reported the status of translations including sentence-level error rates, word-level error rates and also confusion pairs for erroneous sentences.

For the WER calculation, the SCLITE scoring method first set an alignment [19] of the hypotheses (the translated sentences) and the reference texts and then perform a minimum Levenshtein distance function which weights to compute the cost of correct words, insertions (I), deletions (D), substitutions (S) and the number of words in the reference (N). The formula for WER can be stated as Eq. 7.

To know the counts of I, D, C and S for the translated Myanmar sentence "သူမ ဒီမှာ မ နေ တော့ဘူးလား ။" ("She doesn't live in here, right?" in English), firstly, the translated texts are compared to the reference texts, and then calculated the counts as follow:

Scores: (#C #S #D #I) 4 2 0 0
REF: သူမ ဒီမှာ မ နေ တော့ဘူးလား ။
HYP: သူမ ဒီမှာ မ ရှိ တော့ဘူး ။
Eval: S S

For this example, there is no deletions ($D = 0$) or insertions ($I = 0$) and only two substitutions (နေ => ရှိ), (တော့ဘူးလား => တော့ဘူး) are happened and thus the number of correct word C is 4. Using Eq. 7, the SCLITE program calculated the WER value for the given example as 16.67%.

The following example is the evaluation result for Chin-to-Myanmar translation by PBSMT.

Scores: (#C #S #D #I) 3 3 2 1
REF: ***** ခင်ဗျားကို ကျွန်တော်တို့ စောင့် မှာ မ ဟုတ် ဘူး ။
HYP: ငါတို့ မင်း ကို စောင့် မှာ **** ********* မဟုတ်ဘူး ။
Eval: I S S D D S

In this case, one insertion (***** => ငါတို့), three substitutions (ခင်ဗျားကို => မင်း) , (ကျွန်တော်တို့ => ကို) , (ဘူး => မဟုတ်ဘူး) , and two deletions (မ => ***) , (ဟုတ် => *********) are occurred (i.e., $C = 3$, $S = 3$, $D = 2$, $I = 1$) and thus WER is 75%.

The top 10 confusion pairs of PBSMT, HPBSMT and OSM models for both translation directions is shown in Table V.

TABLE V
TOP 10 CONFUSION PAIRS OF PBSTM, HPBSMT, AND OSM FOR MYANMAR-CHIN AND CHIN-MYANMAR TRANSLATIONS

My-Ch (Ref → Hyp)	**Freq**	**Ch-My (Ref → Hyp)**	**Freq**
. → ?	79	ခင်ဗျား → မင်း	23
ka → a	38	ငါ → ကျွန်တော်	14
i → a	30	ပါဘူး → ဘူး	13
nge → ni	21	အဲဒါ → အဲ့ဒါ	11
chu → chuan	20	က → ကို	8
na → em	18	မင်း → ခင်ဗျား	8
hi → chu	16	ငါ့ → ကျွန်တော့်	6
hmeichhia → ani(hmeichhia)	15	မ → ကို	6
nei → awm	14	မင်းရဲ့ → မင်း	6
ang → dawn	13	မဟုတ်ဘူး → ဘူး	6

As a remark based on the detailed analysis of confusion pairs of each model, most confusion pairs are happened because of word segmentation, typing errors, paraphrasing errors and number and text number errors. Some examples of confusion pairs related to word segmentation are (ပါဘူး => ဘူး), (မဟုတ်ဘူး => ဘူး), (ခဲ့ဘူးလား => ဘူးလား), (ဘာတွေ => ဘာ တွေ), (ရမှာလား => မှာလား), and (ခဲ့ဘူး => ဘူး). The confusion pairs (ငါတို့ => ကျွန်တော်တို့), (ကိုယ် => ကျွန်တော်), (မင်း => ခင်ဗျား), (ကျွန်မ => ငါ), (ကျွန်တော် => ကိုယ်), (တို့ => ကျွန်တော်တို့), (စိုးရိမ် => စိတ်ပူ), (ဒါပေမယ့် => ဒါပေမဲ့) are happened because of paraphrasing errors. The next group is occurred according to typing errors such as (ငါ => ငါ့), (အဲဒီ => အဲ့ဒီ), (မေး => မ◌ေး), (သူ့ => သူ), etc. In the other group, errors of number and text number such as (၅ => ငါး), (ကိုး => ၉) are included. To reduce these various groups of confusion pairs, the better pre-processing steps should be applied.

VIII. Conclusion

This paper contributes the first evaluation performance of PBSMT, HPBSMT and OSM approaches from Myanmar-to-Chin (Mizo) and Chin (Mizo)-to-Myanmar. In the experiment, the size of around 15K parallel corpus for Myanmar-Chin language pair is developed. The 10-fold cross validation is also applied. The experimental results based on the scores of the three translation approaches show that OSM approach can obtain not only the highest BLEU scores in both directions but also RIBES and chrF scores in Chin-Myanmar direction, HPBSMT approach is better than the others for Myanmar-Chin, and OSM approach is the best score for Chin-Myanmar language pair. Based on the experimental results, the error analysis and confusion pairs of machine translation between Myanmar and Chin languages are also described.

We have a future plan to perform neural machine translation models for under-resourced langauges such as Myanmar-Chin and Myanmar-Shan.

References

[1] Philipp Koehn, Franz J. Och, Daniel Marcu,"Statistical Phrase-Based Translation", In Proc. of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, pp. 127-133, 2003.

[2] Philipp Koehn,"Tutorial of Statistical machine translation: the basic, the novel, and the speculative", the 11th Conference of the European Chapter of the Association for Computational Linguistics, 2006.

[3] Benjamin Marie and Atsushi Fujita, "Phrase Table Induction Using Monolingual Data for Low-Resource Statistical Machine Translation", ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), vol. 17, issue 3, pp. 1-25, 2008.

[4] Ye Kyaw Thu, Andrew Finch, Win Pa Pa, and Eiichiro Sumita, "A Large-scale Study of Statistical Machine Translation Methods for Myanmar Language", In Proc. of SNLP2016, February 10-12, 2016, Phranakhon Si Ayutthaya, Thailand.

[5] Win Pa Pa,Ye Kyaw Thu, Andrew Finch, Eiichiro Sumita, "A study of statistical machine translation methods for under resourced languages", 5th Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU Workshop), pp. 250-257, 2016.

[6] Ye Kyaw Thu, Vichet Chea, Andrew Finch, Masao Utiyama and Eiichiro Sumita, "A Large-scale Study of Statistical Machine Translation Methods for Khmer Language", 29th Pacific Asia Conference on Language, Information and Computation, Shanghai, China, october 30 - November 1, 2015, pp. 259-269.

[7] Thazin Myint Oo, Ye Kyaw Thu, and Khin Mar Soe, "Statistical machine translation between Myanmar (Burmese) and Rakhine (Arakanese)", In Proc. of ICCA 2018, pp. 36-45, 2018.

[8] Thet Thet Zin, Khin Mar Soe, and Ni Lar Thein, "Myanmar phrases translation model with morphological analysis for statistical Myanmar to English translation system", In Proc. of the 25th Pacific Asia Conference on Language, Information and Computation (PACLIC 2011).

[9] David Chiang, "Hierarchical phrase-based translation", Computational Linguistics, vol. 33, issue 2, pp. 201-228, 2007.

[10] Braune, Fabienne, Anita Gojun, and Alexander Fraser, "Long-distance reordering during search for hierarchical phrase-based SMT", In Proc. of the 16th Annual Conference of the European Association for Machine Translation (EAMT), Trento, Italy, pp. 177-184, 2012.

[11] Amarnath Pathak, Partha Pakray, Jereemi Bentham, "English-Mizo machine translation using neural and statistical approaches", Neural Computing and Applications, pp. 7615-7613, 2013.

[12] Franz Josef Och, and Hermann Ney, "Improved statistical alignment models", In Proc. of the 38th Annual Meeting on Association for Computational Linguistics, pp. 440-447, 2000.

[13] Goutam Majumder, Partha Pakray, Zoramdinthara Khiangte, Alexander Gelbukh,"Multi-word expressions (MWE) for Mizo language: literature survey,", In Proc. of International conference on intelligent text processing and computational linguistics, pp. 623-635, 2016.

[14] Nadir Durrani, Helmut Schmid, Alexander M. Fraser, Philipp Koehn and Hinrich Schutze, "The Operation Sequence Model Combining N-Gram-Based and Phrase-Based Statistical Machine Translation", Computational Linguistics, vol.41, No.2, pp. 185-214, 2015.

[15] https://www.statmt.org/GIZA++.

[16] Christoph Tillmann, "A unigram orientation model for statistical machine translation", In Proc. of HLT-NAACL 2004: Short Papers, pp. 101–104, 2004.

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

[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, pp.177-180, 2007.

[19] Franz Josef Och, and Hermann Ney, "Improved statistical alignment models", in Proc. of the 38th Annual Meeting on Association for Computational Linguistics, pp. 440-447, 2000.

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

[21] (NIST) The National Institute of Standards and Technology, Speech recognition scoring toolkit (sctk), version: 2.4.10, 2015.

[22] Franz Josef Och, "Minimum error rate training in statistical machine translation", in Proc. of the 41st Annual Meeting and Association for Computational Linguistics – vol. 1,Association for Computer Linguistics, Sapporo, Japan, July, pp.160-167, 2003.

[23] Sonja Nießen, Franz Josef Och, Gregor Leusch, Hermann Ney, "An evaluation tool for machine translation: fast evaluation for MT research,", In Proc. of Second International Conference on Language Resources and Evaluation (LREC), pp. 39-45,2000.

[24] Kishore Papineni, Salim Roukos, Todd Ward, Wei Jing Zhu, "BLEU: a method for automatic evaluation of machine translation", in Proc. of 40th Annual meeting of the ACL, pp. 311-318, 2002.

[25] Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, Hajime Tsukada, "Automatic evaluation of translation quality for distant language pairs", In Proc. of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 944-952, 2010.

[26] Ahmed Ali, Steve Renals, "Word Error Rate Estimation for Speech Recognition: e-WER", In Proc. of the 56th Annual Meeting of the Association for Computational Linguistics, vol. 2: Short Papers, pp. 20-24, July 2018, Melbourne, Australia.

[27] Word Error Rate:https://en.wikipedia.org/wiki/Word_error_rate.

[28] ChrF++: http://www.statmt.org/wmt17/pdf/WMT70.pdf

[29] Maja Popović,"CHRF: character n-gram F-score for automatic MT evaluation", In Proc. of the Tenth Workshop on Statistical Machine Translation, pages 392–395, Lisboa, Portugal, 17-18 September 2015.

[30] Liling Tan, Jon Dehdari, Josef van Genabith,"An Awkward Disparity between BLEU / RIBES Scores andHuman Judgements in Machine Translation",In Proc. of the 2nd Workshop on Asian Translation (WAT2015), pp 74 81, Kyoto, Japan, 16th October 2015.