



An empirical analysis on statistical and neural machine translation system for English to Mizo language

Chanambam Sveta Devi¹ · Bipul Syam Purkayastha¹

Received: 26 March 2023 / Accepted: 28 August 2023 / Published online: 13 September 2023

© The Author(s), under exclusive licence to Bharati Vidyapeeth's Institute of Computer Applications and Management 2023

Abstract Machine Translation Systems for under-resource languages encounter quality and comprehension issues. Our research work focuses on the Statistical and Neural approaches methodologies for translating English into Mizo in a specific domain. We created an English-to-Mizo parallel dataset from the National Platform of Language Technology (NPLT) domains, the Bible and other domains as part of the system development. The performance of translations produced by Phrase-Based Statistical Machine Translation (PB-SMT) and Neural Machine Translation (NMT) systems were trained and tested in under-resource and domain-specific circumstances which were then explored thoroughly utilizing automatic and subjective evaluation approaches. The experiment conducted with PB-SMT displayed better results as compared to the state-of-the-art NMT on English to Mizo translation works. The testing quality of our system was evaluated through a suitable example with automatic BLEU and Manual evaluation consisting of two parameters namely, adequacy and fluency.

Keywords PB-SMT · NMT · Machine translation · SentencePiece · Under-resource

1 Introduction

Machine Translation is the study of spontaneously translating languages from one to another using a Machine. Due to an increase in computational power, such as hardware and software technologies, Machine Translation research has yielded promising results for most prominent languages. While there have been successes for languages with abundant resources, relatively fewer scientific studies have been conducted on languages having limited resources, such as the Mizo language. For a machine translation system to produce high-quality translations, it necessitates a substantial amount of parallel corpus that is of better quality [1, 2]. For translation of high quality, it is necessary to possess a comprehensive comprehension of the syntactical and semantic components of both the source and target languages. The significance of researching and creating good MT systems has grown in recent years as a result of rising globalization when individuals from various origins and with varying levels of language proficiency collaborate. Currently, two paradigms are being used to create MT. The first is based on statistical methods, whereas the second is based on artificial neural networks. In the current environment, Machine Translation systems can be categorized as Statistical and Neural Machine Translation. SMT has helped Machine Translation become more widely accepted. It requires the development of SMT models, the input data of which are derived from research done on bilingual training corpus by experienced translation services [3]. The most recent version of Moses Toolkit v4.0, which [4] created for SMT and includes components like Word Alignment, Language Model development, and Phrase Table generation, as some of its supporting elements. SMT has been the subject of numerous studies [5] with promising findings for a variety of language pairs.

Chanambam Sveta Devi and Bipul Syam Purkayastha contributed equally to this work.

✉ Chanambam Sveta Devi
chsveta91@gmail.com

Bipul Syam Purkayastha
bipul_sh@hotmail.com

¹ Department of Computer Science, Assam University, Silchar, Assam 788011, India

Despite its infancy, Neural Machine Translation (NMT) [6] has already shown promising results [1, 7], which has led to a great deal of curiosity and attention. [8], suggested continuous recurrent translation models without the need for alignment or phrasal translation units. However, [9] addressed the subject of rare word occurrence and examined the efficacy of both global and local techniques [10, 11] showing a log-linear framework with SMT characteristics combined with NMT to handle issues such as a lack of vocabulary and poor translation. These architectural characteristics were thoroughly covered in [12]. This approach produces translation even with a sufficient supply of training data, this is noticeably more accurate than SMT [13–15]. And, the traditional NMT models face a common problem, which is the handling of rare or out-of-vocabulary (OOV) words. To tackle this issue, subword-level models, such as SentencePiece, have been proposed. SentencePiece [16] is an unsupervised text tokenizer that aims to divide a sentence into subword units, called pieces, to mitigate the OOV problem. The idea behind SentencePiece is to divide words into smaller subword units and represent them as vocabulary tokens, which enables the model to handle rare and out-of-vocabulary (OOV) words more effectively.

We propose in this paper an NMT model that incorporates SentencePiece for tokenization. The proposed model consists of an encoder-decoder architecture with an attention mechanism. The input and target sentences are first pre-processed with SentencePiece and then fed into the NMT model. Our analysis to the suggested model on several benchmark datasets and compare its performance with the traditional NMT models. The results show that the proposed NMT with SentencePiece significantly outperforms the traditional NMT models in terms of BLEU scores, demonstrating its effectiveness in handling rare and out-of-vocabulary (OOV) words.

The paper evaluated the efficiency of Statistical Machine Translation (SMT) and Neural Machine Translation (NMT) models on English sentences collected from National Platform of Language Technology (NPLT), the bible and other web source. The English sentences are translated manually into Mizo by native speakers to produce a parallel corpus of English–Mizo (en–mz) sentences. This research work looked into the circumstances under which NMT and SMT outperformed one another. Furthermore, they would help us to determine if using specific sentences as training data for MT models affects the quality of the MT output.

It is challenging to develop Statistical Machine Translation (SMT) and Neural Machine Translation (NMT) models for Mizo because Mizo training data are scarce and aren't available. These models require a huge amount of parallel data between languages to create effective machine translation systems. Moreover, parallel corpus is often confined to a certain domains, which results in lower performance when

using machine translation models to translate content outside of the trained domain [17].

SMT and NMT systems were used on total 58,650 sentences of parallel corpus. Table 2. gives the details of the parallel dataset. To validate the observed results for the entire corpus, the automatic evaluation metric BLEU is used [18] and for manual evaluation fluency and adequacy were calculated on the test sentence corpus.

This paper compares the SMT and NMT models for under-resources language pair. Our objective is to gain insights into the ability of en–mz models to translate accurately, and we aim to achieve this by conducting both manual and automatic evaluations of the translations. We used the Automatic Metric i.e., BLEU score, for comparing and analyses the effectiveness of the models, as well as a human evaluation of 100 random sentences. The human evaluation is conducted by two native Mizo speakers who assess the translations provided by all the models under consideration in greater detail.

1.1 Research objective

The Objective of this research is to evaluate and contrast the results generated by a PB-SMT system and an NMT system. The evaluation will be based on linguistic considerations and will be performed using automated metrics for measuring translation quality. In addition, we will separate the two MT systems into their SMT and NMT variants. The goal is divided into two sub-goals.

The first objective compares the BLEU metric was used to evaluate the accuracy of Statistical Machine Translations and Neural machine translations for certain English to Mizo texts.

The second goal resembles the first one, but it focuses on assessing the human aspect. Specifically, we assess and contrast the effectiveness of Statistical and Neural Machine Translation in converting particular English texts to Mizo.

We expect the Machine Translation system based on Phrase-Based Statistics to outperform the Neural Machine Translation system in terms of quality (accuracy), despite of whether it has been trained using parallel corpus from a certain domain.

1.2 Mizo language

Mizo is a member of the Tibeto-Burman language family spoken by approximately 700,000 people primarily in the state of Mizoram, India, as well as Chin State in Burma and the Chittagong Hill Tracts in Bangladesh []. Mizo was previously known as Lushai, Lusei, or Lushei, after the language's most popular dialect, which serves as a lingua franca among the Kuki people. The Mizo language is closely related to the other Tibeto-Burman languages spoken by

these populations. Mizo did not have its own writing script until the arrival of Christian missionaries. Before it was written in the Bengali script. The two missionaries also produced the first Lushai grammar and dictionary, which laid the foundation for the subsequent decades' growth of the Mizo language and literature. The pioneer Welsh Missionaries first used the 25 letters of the Mizo alphabet “a, aw, b, ch, d, e, f, g, ng, h, i, j, k, l, m, n, o, p, r, s, t, ꠘ, u, v, z” which has six “a, aw, e, i, o, u” vowels and 19 consonants “b, ch, d, f, g, ng, h, j, k, l, m, n, p, r, s, t, ꠘ, v, z” in 1984. It is based on roman scripts. Various attempts have been undertaken to manually translate many English-written books into Mizo. The majority of these works have been translated utilising methods of translation such as word-for-word translation, free translation, Semantic translation, Adaptation translation, and Idiomatic translation in which the themes, characters, and stories are retained in the translated texts. Mizo is an agglutinative language with a rich morphological structure [19]. A lexical root is followed by one or more affixes in Mizo words. Person, number, gender, and case markers inflect Mizo words [20]. Mizo language is a tonal language because tone dictates the lexical meaning of words. Mizo has a total of eight tones, four of which are long tones and four of which are short tones. The usage of diacritics in Mizo tonal words is not specified. In terms of Computational Linguistics and Natural Language Processing (NLP), very little research has been conducted on the Mizo language. There is no large-scale reliable parallel corpus available for the Mizo language.

Additionally, the structure of this paper is follows, Sect. 2 explains the previous works, Sect. 3 discusses the detail of corpus collection and the process of preparation. Sect. 4 about the MT system for our experimental setup. Section 5 deals with the training of the MT system, the analysis and evaluation in terms of several measures, and Section 6 conclusions.

2 Previous works

With the introduction of NMT in research on machine translation, scholars have begun to explore the advantages and disadvantages of NMT compared to PB-SMT. This section highlights some studies that have compared the two methods across different scenarios. Although our main focus is on analyzing translations produced by both PB-SMT and NMT in situations with limited resources, we also present summary of studies that have compared the two methods in settings with abundant resources.

After studying various language pairs to compare the effect of two translation systems on the resulting translations. English and Khasi [21] found the PB-SMT better than the NMT in under-resource scenarios. Another study,

mentioned in [22], quantitatively analyzed both SMT and NMT systems using the same dataset to investigate the users' perception and utilization of these systems for translation purposes. Another study, cited as [23], qualitatively compared NMT and SMT for English–Hindi and English–Bengali languages. After studying various language pairs to compare language pairs. The conclusion was that NMT performs better than SMT for simple sentences, but SMT performs better than NMT for translations of all types of sentences. The article referenced as [24] conducted bidirectional machine translation (MT) between English and Myanmar, using both NMT and SMT with pre-ordering included. The results indicate that NMT performs better for English-to-Myanmar translation, whereas SMT outperforms NMT for Myanmar-to-English translation. Another study, mentioned as [25], focused on evaluating the SMT and NMT output using special translators instead of automatic metrics. The findings reveal that NMT can produce more accurate paraphrases with fewer errors, but these errors that do occur are difficult to detect. In contrast, SMT provides translations with more errors, but these errors are easier to identify. These are a few examples of research that compare the two approaches using the BLEU score.

Nevertheless, BLEU produced results that contradicted human judgment [26]. As a result, BLEU alone cannot meet the performance requirements of the evaluation system. To demonstrate the some of we present work that contradicts human judgment with that of BLEU.

As Mizo is a under-resource language, so there are very few resources available for this language. SMT, which dominated the area of MT, and NMT, the present technology that has surpassed SMT, are both significantly reliant on the quality and amount of datasets. Aside from a dearth of data, the language pair under study, English–Mizo, has significant structural, morphological, and phonological variation.

A few applications of Machine Translation were developed in the Mizo language and are still in its early stages. In [27] states that experimenting with the pre-processing steps on the bible domain and testing on 6 pairs of English and Mizo parallel corpus. And showed the Splitting of sentences, Tokenization, True-casing, and Cleaning of this system. For further works, an increase of corpus for the future work of this paper and implementation for better results and performing works of SMT.

In [28] NMT system was educated for translating English to Mizo, using a parallel corpus of 10,675 sentences. Its effectiveness was tested on a separate dataset of 100 sentences, and the results showed that the system was satisfactory in terms of fluency, but not accuracy [29]. The study was to evaluate the performance of the same NMT system in various domains using multiple test datasets. In a study on English to Mizo machine translation systems [30], utilized a training dataset sourced from different online platforms to

compare the effectiveness of SMT and NMT systems. They expanded upon the work of [31] by adding a supplementary training dataset comprising 31,764 parallel sentences. Their models are evaluated using three separate test datasets consisting of 798, 100, and 100 sentences. The models were trained using PB-SMT, as well as NMT methods like LSTM, BiLSTM, and Transformer. The results indicated that the NMT-Transformer model outperformed the baseline system. In Ref. [32] used a transformer as well as an Attention-Based LSTM. With 30,800 datasets of the English–Mizo bible corpus [33], proposed an experimental test on the English–Mizo Statistical Machine Translation with the Bible corpus. The system was analyzed using the automatic scoring methodologies BLEU and METEOR score, as well as manually evaluated by linguistic experts. SMT systems with BLEU scores of 18.71 for English to Mizo and score of 19.44 for Mizo to English perform better than other MT systems when trained with the Language Model’s 5-gram order. The outcomes of the automatic evaluation demonstrate that the MT system performs better as the n-gram order of the LM increases. Despite researchers investigating the English–Mizo MT system.

3 Machine translation system for our experiments set up

In Our Experimental models, we choose SMT systems and another for teaching distinct NMT systems. All configurations are explained below.

3.1 Phrase-based statistical machine translation (PB-SMT)

PB-SMT approaches the machine translation problem by assuming that source and target sentences are translations with a certain probability. PB-SMT uses the probability distribution $p(t|e)$ to predict the target language t , given a source language e . The conditions distribution is calculated based on using the Bayesian model such as $p(t|e)ap(e|t)p(t)$, where $P(e|t)$ is the probability of target t is the translation of source e and $P(t)$ represent the language model which is the probability of target word and t is the good translation given by the equation below:

$$t = \operatorname{argmax}_t p(t|e) = \operatorname{argmax}_e p(e|t)p(t) \tag{1}$$

After the parameters are tuned, the weighted model is log-linear, as shown by Eq. 2.

$$p(x) = \exp \sum_{i=1}^n \lambda_i h_i(x) \tag{2}$$

where \mathbf{n} stands for the total number of feature functions, λ_i for weight, and h_i for feature functions such as language models, translation, and reordering.

3.2 Neural machine translation (NMT)

Neural machine translation (NMT) is a method of machine translation that employs neural networks to determine the probability of a sequence of words, generally moderating full phrases in a single integrated model. NMT model are based on the conditional probability of a word sequence between the source and the target. While the SMT system calculates conditional probability using the Markov assumption, the NMT system learns joint probability with no assumptions. Encoder-Decoder design is typically used in NMT [1, 10, 33]. A basic-based model [1] computes a sequence of outputs $\mathbf{y} = (y_1, \dots, y_T)$ for a given input sequence $\mathbf{x} = (x_1, \dots, x_T)$.

$$y = \operatorname{argmax}_y p(y|x) \tag{3}$$

Encoder reads input sequence x into vector sequence c as an initial step, using Eqs. 4 and 5.

$$h_t = f(x_t, h_{t-1}) \tag{4}$$

and

$$c = q(h_1, \dots, h_{T_x}) \tag{5}$$

Where \mathbf{c} is the context vector created by the hidden states, h_t is the hidden state at time t , \mathbf{f} and \mathbf{q} are non-linear functions, and \mathbf{f} can either be an LSTM or a GRU. Given context vector \mathbf{c} and all previously generated word sequences $(y_1, \dots, y_{t'-1})$ the decoder guesses the following word $\mathbf{y}t'$. The probability of the translation sequence \mathbf{y} is calculated using the equation shown below:

$$p(\mathbf{y}) = \prod_{t=1}^T p(y_t) (\{y_1, \dots, y_{t-1}\}, c) \tag{6}$$

SentencePiece: This technique is helpful for sub-word tokenization [32] for pre-processing text data for machine translation. It is intended to be language-independent and can work with a wide range of languages and scripts. SentencePiece works by encoding text into subword units such as subwords, words, or characters. The size of the subword units, the test coverage, and the informativeness of the subwords are all balanced in this encoding.

4 Corpus collection and preparation

The Statistical and Neural technique is more commonly used in the research of Machine Translation since it requires far fewer computing language resources than the rule-based approach. Nevertheless, parallel corpora of the source and target languages are heavily utilised in the Statistical and Neural approach.

We have used a variety of methods to collect monolingual corpora for the selected language in English. The information gathered relates to religion and others. The religious domain includes the Holy bible [34] and other different domains like tourism and health of Monolingual text corpora that are collected from the National Platform of Language Technology (NPLT) [35] which are freely available on the website and additional short datasets were collected from local blog sites.

Machine Translation systems depend on the bilingual parallel corpora utilised on SMT and NMT training system. The structure and size of a corpus determine the accuracy and efficiency of both systems. A bilingual parallel corpus is a collection of text files in one natural language (source) with their corresponding translated language (target). In this paper, we considered that the aligned text of the English language is translated into the sentences of the Mizo language. This section describes the process of the construction of a corpus. All the collected monolingual corpus is constructed by manually translating through a linguistic person. This text collection is pre-processed further. All the English and Mizo sentences are compiled together into an Excel file and again were split randomly into three subset files using Sklearn toolkits (Train, Tune, and Test file). The pre-processing process includes tokenization, true-casing, and cleaning of the parallel corpus. *Tokenization*: the process of breaking it up into pieces, known as tokens, with the sequence of characters and a specified document unit. The tokens means words, punctuation marks, and numbers in this case. *True-casing*: the process of converting the upper case letter into a lower case letter. It will also reduce data sparsity. *Cleaning*: Long sentences (more than 80 tokens) were eliminated. For this process, we used Moses tokenizer scripts. Table 1, and 2 shows the Monolingual dataset before pre-processing and the statistics of parallel text after pre-processing and gave the total dataset included in pre-processing. The parallel (en–mz) corpus is to be split using Python script into the training, tuning, and testing files.

Table 1 Statistics of the Monolingual dataset

Corpus	# of lines	# of tokens
Monolingual	60,254	2,011,200

Table 2 Statistics of the parallel dataset

Dataset	# on lines	# of tokens
Training (en)	54,150	1,238,000
Training(mz)		1,345,456
Tuning (en)	2500	64,389
Tuning (mz)		69,984
Testing(en)	2000	48,929
Testing (mz)		53,987

5 Experimental setup

Moses [4] toolkit developed a Phrase-Based Statistical Machine Translation System of English to Mizo language. The Phrase Table, the phrases are extracted using the phrase table that was created during training. Retaining the Moses settings: grow-diag-final-and heuristics for word alignment, “msd-bidirectional-fe” for reordering model, and 5-geam language model (LM) with modified Kneser–Ney smoothing [36] using KENLM [37]. Even so, the order of LM does not affect the phrase table. GIZA++ [38] was used for the word alignment toolkit and MERT [39] was used to perform the tuning with batch-MIRA [40] in the translation model for word alignment. The automatic evaluation scoring is done with the help of BLEU.

On the other hand, our NMT systems are built with the OpenNMT [41] toolkits, which are publicly available. This typical NMT solution employs an attentional encoder-decoder network [1]. For 13 epochs, we train a 2-layer LSTM with 500 hidden layers. Using similar training data as the SMT system for comparison purposes (Table 2). The total vocabulary size is 50,002 (English) and 47,301 (Mizo). It is worth noting that another module is also applied to the NMT system’s output. In addition to this system, we run a few exploratory tests to see how changing settings or utilising other methods affects an English–Mizo NMT system.

NMT + SentencePiece (Sp): We apply the SentencePiece model with 32,000 subword units and train the model for English to Mizo Machine Translation with the same scripts available [36]. The Sentencepiece is based on Byte-Pair-Encoding, which is trained in the system with the same process as OpenNMT.

5.1 Experimental result and analysis

5.1.1 Comparison of MT Accuracy

Three different translation models were developed. The SMT baseline system, which is built with default parameters, is denoted as the SMT. NMT stands for the OpenNMT baseline. Moreover, the OpenNMT-based SentencePiece model is constructed and referred to as NMT(Sp). The SMT

and NMT were assessed using a separate test set of 2000 sentences. In analysing the outcomes, the BP denotes the brevity penalty, which takes into account the length of the hypotheses and reference translations. Specifically, HypLen represents the length of the hypotheses, whereas RefLen represents the length of the reference translations (Table 3).

The NMT system has performance poorly, which can be partly attributable to the limited resources made available for its development. Despite the fact that each English sentence can be translated into Mizo in a variety of ways, the system’s effectiveness is also automatically assessed based on a single reference test sentence. We’ll perform a subjective study of the translated outputs to better understand the translation quality; this will be covered in the section that follows. The Baseline SMT system performs better than its NMT competitors in terms of automated scoring, as measured by the BLEU score.

Analysis of BLEU Results based on the Sentence length: The results of our analysis of the Baseline SMT in terms of sentence length for the English to Mizo MT system are displayed in Fig. 1. Our models are categorized as SMT, NMT, and NMT(Sp), which correspond to Statistical and Neural Machine Translation, respectively.

According to the outcomes depicted in Fig. 1, for sentences that are shorter than 10 words in English to Mizo translation, the NMT system performs notably better than other SMT systems. In the case of sentences that are 50 words or more in length, the NMT(Sp) system shows the best performance. Nonetheless, in general, the SMT system is observed to have better performance than both the NMT and NMT(Sp) systems.

Based on the experimental outcomes mentioned above, we can notice a notable difference in the effectiveness of the Baseline SMT system in comparison to NMT systems. Specifically, the effectiveness of the Baseline SMT is considerably different for sentences that are shorter than ten words and sentences that are longer than or equal to ten words. Moreover, it is worth mentioning that the Baseline exhibits greater robustness in handling short sentences compare to the NMT and NMT(Sp) systems.

Based on Human Evaluation: According to [42], the BLEU score is insufficient to evaluate the effectiveness of the method. Even if they are of the same language pair, the results of various studies cannot be compared. As a result, human judgement becomes an important factor in

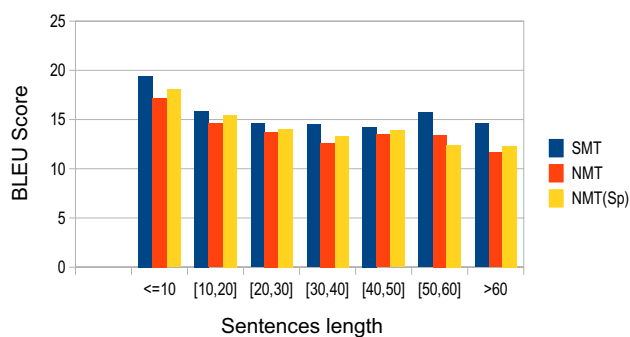


Fig. 1 Evaluation of English to Mizo MT Systems

determining the quality of Machine Translation. Linguistic experts who are native speakers of Mizo and have a good command of the English language evaluates the output of MT quality using two criteria: adequacy and fluency. Adequacy means the amount of meaning fully translated sentences from the reference sentences that are included in the candidate sentences. According to [43], adequacy is “how much of the target translation’s meaning matches that of the gold-standard translation or source”. The evaluator needs to speak both languages well. On a scale from 1 to 5, the no meaning, little meaning, some meaning, most meaning, and entire meaning is considered adequate. And the criteria for fluency are grammar, spelling, word choice, and style; they do not take into account the source. The scale for measuring fluency is 1 to 5, incomprehensible, disfluent, non-native, good and flawless.

The Adequacy and Fluency scores of our MT systems as determined by native linguistic persons are shown in Table 4. Our findings from the automatic evaluation were augmented by the results of the manual evaluation. Here are some examples of the outputs from our SMT systems.

English to Mizo sample input–output

English: be ye strong therefore, and let not your hands be weak: for your work shall be rewarded.

Reference: nimahsela nangni zawng chak takin awm ula, inthlahdah su u, in thiltih man chu in hmu dawng a ni, " a ti a.

PB-SMT: fapa hotu elizafana; chu semaia a ni a, a unaute nen zahnih leh a;

Table 3 Statistic BLEU results for SMT and NMT

System	BLEU	BP	HyfLen	RefLen
SMT	16.23	0.944	49,922	53,998
NMT	15.74	0.933	50,388	53,998
NMT(Sp)	15.82	0.934	49,052	53,998

Table 4 A comparison of the models’ adequacy and fluency based on the sample inputs and respective outputs, with a ranking from 1 to 5

Model	Adequacy	Fluency
Baseline (SMT)	3.6	3.8
NMT	3.0	3.2
NMT(Sp)	3.5	3.3

NMT: chutichuan, nangni mi chak takte u, in kutte hi chak lo ni suh se, in thiltihin lawmman a nei dawn si a.

NMT(Sp): chuvângin, awm hle hle ula, in kutte lo chak suh u, in hna chu a chak dawn si a.

6 Conclusion

In conclusion, we share the initial results of our investigation into how well Statistical and Neural Machine Translation systems perform when translating English to Mizo, a under-resource language, within a specific domain. It's worth mentioning that although the SMT systems had a low BLEU score, they might offer better accuracy for longer sentences compared to the NMT counterparts.

Once more, the SMT system's inflexible alignment approach for longer words or entire sentences highlighted another weakness, namely, that the accuracy and smoothness of the translations were lacking. Based on the human evaluation, the SMT translations were generally smoother than their NMT counterparts because NMT systems require more resources to operate efficiently. A variety of difficulties encountered by us in constructing these machine translation systems were discovered as a result of this research. The tested systems and their outcomes did not meet expectations due to a lack of data and inadequate vocabulary and dictionaries.

This preliminary investigation indicates that the BLEU score has a strong correlation with the human evaluation of this particular language combination. Moving forward, we intend to investigate a larger dataset for this language pair and examine various other machine translation techniques that could enhance the translation quality.

Acknowledgements We are grateful to Ms K. Vanlalruati, Guest Faculty, Mizo Department, Pachuhunga University College, College Veng, Aizawl, Mizoram. and Vanlalhlani Khawhrling, Assam Downtown University, Guwahati, Assam for their support in developing the corpus and for manual evaluation of the corpus and the translation.

Data availability Data can be provides under special request.

Declarations

Competing Interest The authors declare that they have no known completing financial interests or personal relationships that could have appeared to influence the work reported in the paper.

Conflict of Interest The authors declare no conflict of interest.

References

- Mahata S, Das D, Pal S (2016) Wmt2016: a hybrid approach to bilingual document alignment. In: Proceedings of the first conference on machine translation: volume 2, Shared Task Papers, pp. 724–727
- Mahata S, Das D, Bandyopadhyay S (2017) Bucc2017: a hybrid approach for identifying parallel sentences in comparable corpora. In: Proceedings of the 10th workshop on building and using comparable Corpora, pp. 56–59
- Hutchins J (1995) Machine translation: a brief history, concise history of the language sciences: from the sumerians to the cognitivists. edited by efk koerner and re asher. Oxford: Pergamon Press (431–445). <https://www.infoamerica.org/documentospdf/bar05.pdf> on 10(08), 2021 (1995)
- Koehn P, Hoang H, Birch A, Callison-Burch C, Federico M, Bertoldi N, Cowan B, Shen W, Moran C, Zens R et al. (2007) Moses: open source toolkit for statistical machine translation. In: Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions, pp. 177–180
- Koehn P (2009) Statistical machine translation. Cambridge University Press, Cambridge
- Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. arXiv preprint [arXiv:1409.0473](https://arxiv.org/abs/1409.0473)
- Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, Krikun M, Cao Y, Gao Q, Macherey K et al. (2016) Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint [arXiv:1609.08144](https://arxiv.org/abs/1609.08144)
- Kalchbrenner N, Blunsom P (2013) Recurrent continuous translation models. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 1700–1709
- Luong M-T, Sutskever I, Le QV, Vinyals O, Zaremba W (2014) Addressing the rare word problem in neural machine translation. arXiv preprint [arXiv:1410.8206](https://arxiv.org/abs/1410.8206)
- Luong M-T, Pham H, Manning CD (2015) Effective approaches to attention-based neural machine translation. arXiv preprint [arXiv:1508.04025](https://arxiv.org/abs/1508.04025)
- He W, He Z, Wu H, Wang H (2016) Improved neural machine translation with smt features. In: Proceedings of the AAAI conference on artificial intelligence, vol. 30
- Cho K, Van Merriënboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint [arXiv:1409.1259](https://arxiv.org/abs/1409.1259)
- Doherty S, O'Brien S, Carl M (2010) Eye tracking as an mt evaluation technique. Mach Transl 24:1–13
- Vaswani A, Zhao Y, Fossium V, Chiang D (2013) Decoding with large-scale neural language models improves translation. In: Proceedings of the 2013 conference on empirical methods in natural language processing, pp. 1387–1392
- Liu S, Yang N, Li M, Zhou M (2014) A recursive recurrent neural network for statistical machine translation. In: Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: long papers), pp. 1491–1500
- Kudo T, Richardson J (2018) Sentencepiece: a simple and language independent subword tokenizer and detokenizer for neural text processing. arXiv preprint [arXiv:1808.06226](https://arxiv.org/abs/1808.06226)
- Koehn P, Knowles R (2017) Six challenges for neural machine translation. ArXiv preprint [arXiv:1706.03872](https://arxiv.org/abs/1706.03872)
- Papineni K, Roukos S, Ward T, Zhu W-J (2002) Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th annual meeting of the association for computational linguistics, pp. 311–318
- Chhange L, et al. (1989) The grammar of simple clauses in mizo. In: Papers in southeast asian linguistics no. 11: Southeast Asian Syntax. Pacific Linguistics.
- Fanai LT (2015) Tones in Mizo language. J Human Soc Sci 1(1)

21. Singh TD, Hujon AV (2020) Low resource and domain specific english to khasi smt and nmt systems. In: 2020 international conference on computational performance evaluation (ComPE), pp. 733–737. IEEE
22. Esperança-Rodier E, Rossi C, Bèrard A, Besacier L (2017) Evaluation of nmt and smt systems: a study on uses and perceptions. In: 39th conference translating and the computer
23. Mahata SK, Mandal S, Das D, Bandyopadhyay S (2018) Smt vs nmt: a comparison over Hindi & Bengali simple sentences. arXiv preprint [arXiv:1812.04898](https://arxiv.org/abs/1812.04898)
24. Wang R, Ding C, Utiyama M, Sumita E (2018) English-Myanmar nmt and smt with pre-ordering: Nict's machine translation systems at wat-2018. In: Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation: 5th Workshop on Asian Translation: 5th Workshop on Asian Translation
25. Mutal J, Volkart L, Bouillon P, Girletti S, Estrella P (2019) Differences between smt and nmt output—a translators' point of view. In: Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019), pp. 75–81
26. Jassem K, Dwojak T (2019) Statistical versus neural machine translation—a case study for a medium size domain-specific bilingual corpus. *Poznan Stud Contemporary Linguistics* 55(2):491–515
27. Devi CS, Purkayastha BS (2020) Steps of pre-processing for english to mizo smt system. In: Machine learning, image processing, network security and data sciences: second international conference, MIND 2020, Silchar, India, July 30–31, 2020, Proceedings, Part II 2, pp. 156–167. Springer
28. Thihlum Z, Khenglawt V, Debnath S (2020) Machine translation of English language to Mizo language. In: 2020 IEEE international conference on cloud computing in emerging markets (CEEM), pp. 92–97. IEEE
29. Pathak A, Pakray P, Bentham J (2019) English–mizo machine translation using neural and statistical approaches. *Neural Comput Appl* 31(11):7615–7631
30. Lalrempuii C, Soni B, Pakray P (2021) An improved english-to-mizo neural machine translation. *Trans Asian Low-Resource Lang Inf Process* 20(4):1–21
31. Lalrempuii C, Soni B (2020) Attention-based English to Mizo neural machine translation. In: Machine learning, image processing, network security and data sciences: second international conference, MIND 2020, Silchar, India, July 30–31, 2020, Proceedings, Part II 2, pp. 193–203. Springer
32. Devi CS, Purkayastha BS, Meetei LS (2022) An empirical study on English–Mizo statistical machine translation with bible corpus. *Int J Electric Comput Eng Syst* 13(9):759–765
33. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. *Adv Neural Inf Process Syst* 27
34. GNB:Bible You Version Homepage. <https://www.bible.com/en-GB/bible/2163/>
35. NLPT. <https://nplt.in/demo/>
36. Kneser R, Ney H (1995) Improved backing-off for m-gram language modeling. In: 1995 international conference on acoustics, speech, and signal processing, vol. 1, pp. 181–184. IEEE
37. Heafield K (2011) Kenlm: Faster and smaller language model queries. In: Proceedings of the sixth workshop on statistical machine translation, pp. 187–197 (2011)
38. Casacuberta F, Vidal E (2007) Giza++: training of statistical translation models. Retrieved October 29, 2019
39. Och FJ, Ney H (2003) A systematic comparison of various statistical alignment models. *Comput Linguist* 29(1):19–51
40. Cherry C, Foster G (2012) Batch tuning strategies for statistical machine translation. In: Proceedings of the 2012 conference of the North American Chapter of the association for computational linguistics: human language technologies, pp. 427–436
41. Klein G, Kim Y, Deng Y, Senellart J, Rush A (2017) OpenNMT: Open-source toolkit for neural machine translation. In: Proceedings of ACL 2017, system demonstrations, pp. 67–72. Association for Computational Linguistics, Vancouver, Canada. <https://www.aclweb.org/anthology/P17-4012>
42. Human Evaluation of Machine Translation. <https://tech.ebayinc.com/engineering/human-evaluation-of-machine-translation/>
43. Post M (2018) A call for clarity in reporting bleu scores. arXiv preprint [arXiv:1804.08771](https://arxiv.org/abs/1804.08771)

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.