Neural vs Statistical Machine Translation: Revisiting the Bangla-English Language Pair

4 authors:

Md. Arid Hasan
Daffodil International University
23 PUBLICATIONS 47 CITATIONS

Firoj Alam
Qatar Computing Research Institute
130 PUBLICATIONS 1,660 CITATIONS

Shammur Absar Chowdhury
Qatar Computing Research Institute
74 PUBLICATIONS 798 CITATIONS

Naira Khan
Institute of Education and Research
13 PUBLICATIONS 59 CITATIONS

Neural *vs* Statistical Machine Translation: Revisiting the Bangla-English Language Pair

Md. Arid Hasan
Cognitive Insight Limited, Bangladesh
arid.hasan.h@gmail.com

Firoj Alam
QCRI, Qatar
fialam@hbku.edu.qa

Shammur Absar Chowdhury
QCRI, Qatar
shchowdhury@hbku.edu.qa

Naira Khan
Dhaka University, Bangladesh
nairakhan@du.ac.bd

***Abstract*—Machine translation systems facilitate our communication and access to information, taking down language barriers. It is a well-researched area of Natural Language Processing (NLP), especially for resource-rich languages (e.g., language pairs in Europarl Parallel corpus). Besides these languages, there is also work on other language pairs including the Bangla-English language pair. In the current study, we aim to revisit both Statistical Machine Translation (SMT) and Neural Machine Translation (NMT) approaches using well-known, publicly available corpora for the Bangla-English (Bangla to English) language pair. We reported how the performance of the models differ based on the data and modeling techniques; consequently, we also compared the results obtained with Google's machine translation system. Our findings, across different corpora, indicates that NMT based approaches outperform SMT systems. Our results also outperform existing baselines by a large margin.**

***Index Terms*—Machine Translation, Bangla-to-English, Statistical Machine Translation, Neural Machine Translation, Bidirectional LSTM**

I. Introduction

With the emergence of neural sequence-to-sequence (seq2seq) [1] models, the task of automated translation from one language to another has undergone rapid advancement and is currently one of the most substantial fields in the NLP community. Despite the recent advancements in the field, the early concept of automated translation systems can be traced back to the 1950s [2] – motivated by translating information from foreign sources during the Cold War. Since then machine translation research has entered a different era, involving rule-based and SMT systems.

The idea of SMT, derived from information theory [3], is heavily dependent on the parameters of statistical models obtained from bilingual text corpora. These traditional data-driven SMT approaches, word- [4] and phrase-based [5] models, dominated the field of machine translation with their performance and made it useful for various applications.

With the recent advancement in neural networks and increased availability of computational resources, Neural Machine Translation (NMT) has gained significant momentum in the 20^{th} century [6]. In the last five years, it has reached state-of-the-art performance levels [7] and large scale deployment has also started. Particularly, the literature with NMT techniques report higher performance for resource-rich languages such as English to German [8] and English to French [9]. Unlike traditional techniques, the advantage of NMT is that it learns the mapping from input to output in an end-to-end fashion. In which it jointly learns the parameters in order to maximize the performance of the translation [10], [11].

While much of the literature reports the success of NMT approaches, some studies have also reported its limitations. The study of Koehn et al. [12] reports the limitations of current NMT systems including out-of-domain data, highly inflected categories lead to low-frequency words, amount of training data, long sentences and issues with word-alignment [12].

Based on the current developments and literature, in this study, we aim to revisit the different MT techniques for the *Bangla to English* (Bangla $\rightarrow$ English) *language pair*. Our contributions include, (1) conducting experiments using SMT and NMT approaches, (2) consolidating data from different sources and evaluating them to understand the effectiveness of these approaches, and (3) present a comparative analysis and provide future directions.

The motivation behind this study is two-faceted: (a) to enrich the Bangla NLP research community - as to the best of our knowledge, this is one of the first study to compare both SMT and NMT approaches for the Bangla to English language pair using publicly[1] available data; and (b) to understand what the necessary steps are and the resources required to outperform current state-of-the-art systems for the Bangla to English language pair.

For the current study, we use a bidirectional LSTM (BiLSTM) network to determine the state-of-the-art performance in comparison with other language pairs. On the other hand, our interest in using SMT for comparison, is due to its success in various MT systems [13], for different language pairs such as English to German, English to French, English to Spanish, etc. As for evaluating the quality of our designed MT models, we reported a BLEU (bilingual evaluation understudy) score.

The structure of this paper is as follows. Section II, provides a brief overview of the existing work on Bangla MT systems. In Section III, we discuss the datasets that we use in this study. We present the approaches that we use for our experiments in Section IV. Following that in Section V, we discuss the details of the experiments and report the results. In Section VI, we

[1] Except for Penn Treebank Bangla-English parallel corpus, which can be accessed by requesting to banglanlp@gmail.com.

978-1-7281-5242-4/19 ©2019 IEEE

discuss the results of the experiments. Finally, we conclude our work in Section VII.

II. Related Work

There have been many endeavors for both Bangla to English and English to Bangla machine translation systems, which include the development of a Bangla-English parallel corpus, evaluation, designing rule-based and automated systems by examining various machine learning algorithms [14]–[18].

In [19], the authors used the Indic Languages Multilingual Parallel Corpus [20] and achieved a BLEU score of 13.98 using Many-to-One phrase-based SMT. The experimental settings consist of a maximum sentence length of 50, a 3-gram language model, and the grow-diag-final and heuristic for extracting phrases [21]. In another study, authors designed a set of rules by analyzing Bangla and English grammars [16]. The rule-based system translated 25 out of 27 sentences correctly and the other two sentences are 75% and 33% correct respectively. In [15], authors report a phrase-based system, which used the EMILLE[2] and KDE4[3] corpora and achieved an overall BLEU score of 11.7 with the use of 8-gram language model, Moses, GIZA++, MERT, and preposition handling.

In [17], the authors used Context-Free-Grammars (CFGs) for English to Bangla machine translation, which uses morphological analysis, syntactic transfer, and sentence construction rules. In another study [14], the authors report active learning strategies for SMT and achieved a maximum BLEU score of $0.057(\sim 5.7\%)$ for Bangla to English machine translation using Hansardsand EMILLE corpora. In [22], authors report LSTM-based seq2seq model with attention mechanism and achieved a BLEU score of 10.92 with an accuracy of 20.5%. The most recent and notable work for the Bangla-English language pair has been done by Mumin et al. [23], where authors proposed a phrase-based SMT based solution, which is experimented and evaluated using one of the largest Bangla-English parallel corpora [24]. The study used log-linear phrase-based SMT, which outperform other phrase-based SMT techniques. The authors also highlight the complexities and the challenging issues that need to be addressed in future research.

Another notable work, for Bangla-English language pair, can be found in [25]. The authors, in the study, investigated both SMT and NMT results on Bangla English language pair. The study reports BLEU score of 16.56 using vanilla phrasal approach with word break, and 20.23 by using NMT approach with synthetic data augmentation for Bangla to English language pairs. Unfortunately, these results, in [25], are not directly comparable to ours, due to difference in dataset and lack of access to it.

Current state-of-the-art performance for resource-rich language pairs are relatively higher compared to the Bangla-English language pair. In [18], authors report an NMT approach, which jointly learns to align and translate using a RNN Encoder-Decoder. The reported BLEU score is 28.45 on English-French parallel corpora with the of maximum sentence length 50.

Our study differs from the reported studies in a way that we compare both SMT and NMT based approaches and present the utility of a sequence to sequence BiLSTM approach. From the evaluation across different corpora we report that our system outperforms existing baselines [19], [23] on the Bangla to English language pair.

III. Data

For the experiment we used corpora that are obtained from different sources, discussed below.

- Indic Languages Multilingual Parallel Corpus (ILMPC): This is publicly available through WAT [20], which has been made available in the Workshop on Asian Translation (WAT). The Indic Languages Multilingual Parallel Corpus consists of 7 parallel languages. The text in the Indic Languages Multilingual Parallel Corpus is collected from OPUS[4]. The ILMPC Bengali-English parallel corpus consists of $\sim$337K parallel sentences in the training set, 500 sentences in the development set and $1,000$ sentences in the test set. More details about these corpora can be found in [20].
- Six Indian Parallel Corpora [26] (SIPC): Six Indian Parallel Corpora consist of six languages. These parallel sentences were collected from the top-100 most-viewed documents from each language's Wikipedia [26]. The training, development and test sets consists of $\sim$20K, 914 and $1K$ parallel sentences, respectively.
- Penn Treebank Bangla-English parallel corpus (PTB): This corpus has been developed by the Bangladesh team of the PAN Localization Project [5], in which source English sentences have been collected from the Penn Treebank corpus. The dataset has been translated by expert translators, which consists of $1,313$ English-Bangla sentence pairs.
- SUPara Corpus [27], [28]: This corpus has been developed by the Shahjalal University of Science and Technology (SUST), in which sentences consist of different genres such as Literature, Journalistic, Instructive, Administrative, and External Communication. The publicly available version (SUPara0.8M) of this corpus contains $\sim$70.5K Bangla-English sentence pairs [27].

IV. Methodology

For our comparative study, for Bangla to English (BN $\rightarrow$ EN) language pair, we chose to use two different approaches of MT: *SMT* and *NMT*. As for SMT experiments, we used Moses statistical MT tool [21] and the NMT experiments are conducted using the OpenNMT toolkit [29]. We ran the initial experiments using training and development data sets and then, we evaluated the system using the test set. The performance was computed using BLEU and NIST scores [30]. Due to limited space, we are not reporting the NIST score in this paper.

[2]Corpus developed by Lancaster University, UK, and the Central Institute of Indian Languages (CIIL), Mysore, India.

[3]The corpus has been developed under OPUS http://opus.nlpl.eu

[4]http://opus.nlpl.eu

[5]http://www.panl10n.net

A. Preprocessing

For the SMT experiment, we tokenized the corpus using the Moses tokenizer. As a part of the tokenization process, all English sentences were transformed into lowercase. We limited our sentence length to 40, and removed longer sentences to reduce computational time for word-alignment process [13].

For NMT experiments, we tokenized the Bangla sentences using our developed Bangla tokenizer and the English sentences were transformed into lowercase. We did not tokenize English sentences. Then, we preprocessed our corpus using the OpenNMT preprocessing script. In OpenNMT preprocessing step, we set our sentence length limit to 50 and removed longer sentences to get better translation quality [12]. We split the sentences into tokens and tokens were then transformed into tensor values. Using these tensor values, we generated training, validation, and vocabulary files to train our NMT models.

B. Training

1) SMT: For the SMT training, we used the SRILM tool for building our Language Model (LM) and the Moses decoder for training. In language modeling, we used google one billion word [31], English monolingual data from ILMPC Corpus, English monolingual sentences from Europarl and News Commentary[6] corpora. Our language model contains approximately 1.5B word tokens. We trained 3-gram and 5-gram language models using the SRILM toolkit [32]. We used GIZA++ for word alignment in order to train a phrase-based SMT model.

2) NMT: For the NMT training, we used the BiLSTMs network-based approach. LSTMs [33] is a form of Recurrent Neural Network (RNN), which is widely used for capturing long-term dependencies. To predict the output of an input sequence RNNs capture all previous information in a memory cell, which is limited in predicting the output from a very long distance. To overcome this limitation, LSTMs were introduced, which consists of input, output and forget gates and are capable of capturing the long-term dependencies. We initialized hidden layer 500, word embedding size for both source and target 500, number of layers 2, and saved our models in every 10,000 steps. During the training, we initialized the model parameters using pre-trained word-embeddings. For English, we used the Glove model [34] and for Bangla we used the word2vec model [35]. The pre-trained embedding dimension is 300 for both the models. To avoid overfitting of the model we used a dropout rate of 0.3.

V. Experiments and Results

To evaluate the performances of the (BN $\rightarrow$ EN) models, we used Moses scripts to calculate the BLEU score [30]. *We conducted two different sets of experiments:*

1) *Exp Setting 1:* In this experiment we use ILMPC, SIPC, and PTB corpora. We mainly use ILMPC corpus for training, and evaluation. In order to improve the performance we merged all the training sets from these datasets into a training set. We used the ILMPC development set and test set in order to optimize the parameters and evaluate the systems, respectively. For the test setting, we removed all the sentences, from the Bangla parallel set, that contained any English words, resulting in 956 sentences instead of 1000 in the ILMPC test set. We have used this dataset, containing a total of $346,845$ parallel sentences (in Table I), to run the experiments using both SMT and NMT based approaches.
2) *Exp Setting 2:* For this experiment we used the SUPara corpus. We maintained the official training (SUPara0.8M), development (SUParadev2018) and test (SUParatest2018) set, which are publicly available [27], [28], as presented in Table I. From our experimental study (Exp Setting 1), we realized that NMT performs better, therefore, we only conducted experiments using NMT with this dataset.

[6]http://www.statmt.org/wmt09/translation-task.html

Table I: Training, development and test data statistics. Bangla (BN), English (EN), Merged - merged dataset for Bangla to English (BN $\rightarrow$ EN) translation.

Data Set	# of Sent	# of Tokens	Source
Train	346,845	2.92M (BN), 2.42(EN)	Merged
Development	500	8,838 (BN), 7,762 (EN)	dev set (ILMPC)
Test	956	15,528 (BN), 13,612 (EN)	test set (ILMPC)
Train	70,861	886,337 (BN), 995,871 (EN)	SUPara0.8M
Development	500	9,157 (BN), 10,817 (EN)	SUParadev2018
Test	500	9,141 (BN), 10,819 (EN)	SUParatest2018

A. Results on Exp Setting 1:

In Table II, we present the performances of SMT and NMT systems including the baseline and Google Translator's results on the test set of our first experimental setting. From the results, we observe that with our combined dataset along with 3- and 5-gram language models our system outperforms baseline results. Comparing the results with the single dataset (row 2-3), with our merged data, we observe that there is a reasonable improvement in performance. From the results, it is also evident that a 5-gram language model is better than a 3-gram one. Note that in our merged dataset experiment, we retrain LM by combining the three datasets mentioned in Section IV-B1. Then, we used the Moses tokenizer to preprocess them. We realized that the tokenizer plays a significant role in performance, therefore, we aim to improve it in our future study.

The NMT results for single and merged datasets are significantly higher. We only report results with the combined dataset due to limited space, which proves the effectiveness of NMT results and it also corroborates current state-of-the-art.

For comparison, we also report the results of Google Translator, which is lower than all the reported results in Table II.

B. Results on Exp Setting 2:

In Table III, we report results on the SUPara test set (SUParatest2018). As mentioned earlier, we only used the SUPara training set (SUPara0.8M) to train the model. From

Table II: Results on the test set for the Bangla to English (BN $\rightarrow$ EN) language pair. Indic Languages Multilingual Parallel Corpus (ILMPC), Penn Treebank Bangla-English parallel corpus (PTB), Six Indian Parallel Corpora (SIPC). EN-Emb: Glove Pretrained Embeddings, BN-Emb: Bangla Pretrained Embeddings.

Training Set	Experiments	BLEU
	google-translator	12.59
	SMT	
ILMPC	Baseline [19]	*14.17*
ILMPC	3-gram LM	13.24
ILMPC	5-gram LM	13.49
ILMPC + SIPC + PTB	3-gram LM	**14.61**
ILMPC + SIPC + PTB	5-gram LM	**14.82**
	NMT	
ILMPC + SIPC + PTB	BiLSTM	**15.24**
ILMPC + SIPC + PTB	BiLSTM + BN-Emb	**15.56**
ILMPC + SIPC + PTB	BiLSTM + BN-Emb, EN-Emb	**15.62**

the results it is evident that we obtain a higher performance gain using different NMT based approaches. Interestingly, Google Translator's results are far better than any other results including ours. We observe that the SUPara dataset (SUPara0.8M) are cleaner and simpler than the ILMPC dataset on which Google Translator performs better.

Table III: NMT results on the SUPara corpus for the Bangla to English (BN $\rightarrow$ EN) language pair. Trained using the SUPara training set (SUPara0.8M) and tested on the SUPara test set (SUParatest2018).

Training Set	Experiments	BLEU
	SMT	
SUPara0.8M	shu-torjoma [23]	*17.43*
	NMT	
SUPara0.8M	BiLSTM	**17.98**
SUPara0.8M	BiLSTM + BN-Emb	**19.28**
SUPara0.8M	BiLSTM + BN-Emb, EN-Emb	**19.76**
	Google Translator	***28.09***

VI. Discussion

To understand the performance of the models, we conducted a detailed error analysis comparing the source (S), automatic (A) and gold (G) translation of the target sentences. Our exploration indicates that one of the most common source of mismatch pairs are due to insertion or deletion. For example, **(S)**: মোরি, তাঁর সাথে আমি কথা বলব। is translated to **(A)**: '*morrie, i'll talk to him.*' instead of **(G)**: '**morrie, i'll go talk to him.**'. Here the word 'go' is deleted from the output. However, it is to be noted that even though the above sentence pairs are said to be mismatched, the semantic meaning remains intact and represent the source correctly.

Unlike the above benign error, there are translated instances that changes the meaning of the source sentences completely, like: (**S**: ভারতের কোন জায়গায় রকেট নিয়ন্ত্রণের ব্যবস্থা প্রস্তুত হয়? , **G**: **which is the place in india that manufactures rockets?**) $\rightarrow$ (**A**: *any place in india is prepared to control the rules of any kind?*). These types of errors are mostly observed in sentences due to the presence of rare words (e.g. রকেট - 'rocket'). While studying the causes, we realized that the datasets contain many rare (low frequency) words specially in the Bangla training set, in comparison to the English one and is one of the main reasons for lower performance of the model. We also observed that the quality of the translation is a big factor in these pairs (see the gold translation, G, in the above example) and an important factor for evaluating the performance.

Along with the aforementioned issues, other challenges for the Bangla-English pair are: i) morphological richness [36], results in highly-inflected words, ii) limited resources (e.g., amount of training data covering different domains) among others including the issues highlighted in [23].

Compared to the resource-rich languages, the difference in performance for the Bangla to English language pair is very high. As seen in [37], the performance of English-French MT using the WMT 2014 English-French dataset with a transformer model has a BLEU score of 41.0. This exhibits that more effort is required for Bangla-English machine translation research.

VII. Conclusions

In this study, we explored machine translation approaches for the Bangla to English (BN $\rightarrow$ EN) language pair. We compared the performance of SMT (phrase-based) vs NMT based approaches, in which we observed that NMT outperforms SMT based techniques. Our preliminary experiments outperform existing baselines on both the ILMPC and the SUPara test sets. Compared to the Google Translator, our system performs better on the ILMPC data set but not on the SUPara dataset. Our study reveals that tokenization is an important step and a well-designed tokenizer is necessary for Bangla. For NMT based techniques, its performance is highly dependent on parameter optimization and architecture, which we also plan to explore in the future.

Acknowledgment

The research leading to the results has been supported and funded by Cognitive Insight Limited – https://cogniinsight.com/.

References

[1] O. Kuchaiev, B. Ginsburg, I. Gitman, V. Lavrukhin, J. Li, H. Nguyen, C. Case, and P. Micikevicius, "Mixed-precision training for nlp and speech recognition with openseq2seq," 2018.

[2] T. Poibeau, *Machine translation*. MIT Press, 2017.

[3] P. Brown, J. Cocke, S. D. Pietra, V. D. Pietra, F. Jelinek, R. Mercer, and P. Roossin, "A statistical approach to language translation," in *Proc. of the 12th conference on Computational linguistics*. Association for Computational Linguistics, 1988, pp. 71–76.

[4] S. Vogel, H. Ney, and C. Tillmann, "Hmm-based word alignment in statistical translation," in *Proc. of the 16th conference on Computational linguistics*. Association for Computational Linguistics, 1996, pp. 836–841.

[5] P. Koehn, *Statistical machine translation*. Cambridge University Press, 2009.

[6] A. Waibel, A. N. Jain, A. E. McNair, H. Saito, A. Hauptmann, and J. Tebelskis, "Janus: A speech-to-speech translation system using connectionist and symbolic processing strategies," in *Proc. of the ICASSP*, 1991, pp. 793–796.

[7] S. Jean, O. Firat, K. Cho, R. Memisevic, and Y. Bengio, "Montreal neural machine translation systems for wmt-15," in *Proc. of the 10th WSMT*. Lisbon, Portugal: Association for Computational Linguistics, September 2015, pp. 134–140. [Online]. Available: http://aclweb.org/anthology/W15-3014

[8] S. Jean, K. Cho, R. Memisevic, and Y. Bengio, "On using very large target vocabulary for neural machine translation," *arXiv preprint arXiv:1412.2007*, 2014.

[9] M.-T. Luong, I. Sutskever, Q. V. Le, O. Vinyals, and W. Zaremba, "Addressing the rare word problem in neural machine translation," *arXiv preprint arXiv:1410.8206*, 2014.

[10] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[11] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," *arXiv preprint arXiv:1409.1259*, 2014.

[12] P. Koehn and R. Knowles, "Six challenges for neural machine translation," *arXiv preprint arXiv:1706.03872*, 2017.

[13] P. Koehn and J. Schroeder, "Experiments in domain adaptation for statistical machine translation," in *Proc. of the 2nd WSMT*, 2007, pp. 224–227.

[14] G. Haffari, M. Roy, and A. Sarkar, "Active learning for statistical phrase-based machine translation," in *Proc. of the NACL*. ACL, 2009, pp. 415–423.

[15] M. Z. Islam, J. Tiedemann, and A. Eisele, "English to bangla phrase-based machine translation," in *Proc. of the 14th EACL*, 2010.

[16] J. Francisca, M. M. Mia, and S. M. Rahman, "Adapting rule based machine translation from english to bangla," *Indian Journal of Computer Science and Engineering (IJCSE)*, vol. 2, no. 3, pp. 334–342, 2011.

[17] S. S. Ashrafi, M. H. Kabir, M. M. Anwar, and A. Noman, "English to bangla machine translation system using context-free grammars," *International Journal of Computer Science Issues (IJCSI)*, vol. 10, no. 3, p. 144, 2013.

[18] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[19] T. Banerjee, A. Kunchukuttan, and P. Bhattacharyya, "Multilingual indian language translation system at wat 2018: Many-to-one phrase-based smt," in *Proc. of the 5th Workshop on Asian Translation*, 2018.

[20] T. Nakazawa, S. Kurohashi, S. Higashiyama, C. Ding, R. Dabre, H. Mino, I. Goto, W. P. Pa, A. Kunchukuttanand, and S. Kurohashi, "Overview of the 5th Workshop on Asian Translation," Tech. Rep., 2018. [Online]. Available: http://www2.nict.go.jp/astrec-att/

[21] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens *et al.*, "Moses: Open source toolkit for statistical machine translation," in *Proc. of the 45th ACL*, 2007, pp. 177–180.

[22] N. F. Liu, J. May, M. Pust, and K. Knight, "Augmenting statistical machine translation with subword translation of out-of-vocabulary words," *arXiv preprint arXiv:1808.05700*, 2018.

[23] M. A. A. Mumin, M. H. Seddiqui, M. Z. Iqbal, and M. J. Islam, "shu-torjoma: An english↔bangla statistical machine translation system," *Journal of Computer Science (Science Publications)*, 2019.

[24] M. A. Al Mumin, A. A. M. Shoeb, M. R. Selim, and M. Z. Iqbal, "Supara: A balanced english-bengali parallel corpus," *SUST Journal of Science and Technology*, pp. 46–51, 2012.

[25] S. Dandapat and W. Lewis, "Training deployable general domain mt for a low resource language pair: English–bangla," 2018.

[26] M. Post, C. Callison-Burch, and M. Osborne, "Constructing parallel corpora for six indian languages via crowdsourcing," in *Proc. of the 7th WSMT*. ACL, 2012, pp. 401–409.

[27] M. A. Al Mumin, M. H. Seddiqui, M. Z. Iqbal, and M. J. Islam, "Supara0.8m: A balanced english-bangla parallel corpus," 2018. [Online]. Available: http://dx.doi.org/10.21227/gz0b-5p24

[28] M. A. Al Mumin, S. M. H., M. Z. Iqbal, and M. J. Islam, "Supara-benchmark: A benchmark dataset for english-bangla machine translation," 2018. [Online]. Available: http://dx.doi.org/10.21227/czes-gs42

[29] G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. M. Rush, "Opennmt: Open-source toolkit for neural machine translation," *arXiv preprint arXiv:1701.02810*, 2017.

[30] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: A method for automatic evaluation of machine translation," in *Proc. of the 40th ACL*, ser. ACL '02. Stroudsburg, PA, USA: ACL, 2002, pp. 311–318. [Online]. Available: https://doi.org/10.3115/1073083.1073135

[31] C. Chelba, T. Mikolov, M. Schuster, Q. Ge, T. Brants, P. Koehn, and T. Robinson, "One billion word benchmark for measuring progress in statistical language modeling," *arXiv preprint arXiv:1312.3005*, 2013.

[32] A. Stolcke, "Srilm-an extensible language modeling toolkit," in *7th ICSLP*, 2002.

[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[34] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. of EMNLP*, 2014, pp. 1532–1543.

[35] F. Alam, S. A. Chowdhury, and S. R. H. Noori, "Bidirectional lstms—crfs networks for bangla pos tagging," in *Proc. of ICCIT*. IEEE, 2016, pp. 377–382.

[36] M. A. Hasan, F. Alam, and S. R. H. Noori, "A collaborative platform to collect data for developing machine translation systems," in *Proc. of International Joint Conference on Computational Intelligence*. Springer, 2020, pp. 407–416.

[37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.