Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 1–8

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Transfer Learning Based Neural Machine Translation of English-Khasi on Low-Resource Settings

Aiusha V Hujon[a,c,*], Thoudam Doren Singh[b], Khwairakpam Amitab[a]

[a] *Department of Information Technology, North Eastern Hill University, Shillong, India*
[b] *Department of Computer Science and Engineering, National Institute of Technology Silchar, India*
[c] *Department of Computer Science, St. Anthony's College, Shillong, India*

Abstract

Machine translation for low-resource language can be improved using various techniques. One such technique is the application of knowledge learned by training a model with high-resource language pair to another model with a low-resource language pair. The paper discusses the experiments and improvement of the results of neural machine translation using transfer learning for the English-Khasi language pair. Long short-term memory is used as the backbone architecture for the transfer learning model. The essential technique is the shared vocabulary, constructed utilizing the subword unit of byte pair encoding of the two pairs of languages and the subword unit of byte pair encoded datasets. Analysis and evaluation of the experimental output using human subjective evaluation, statistical evaluation, and automatic evaluation show positive results for the transfer learning system. A thorough analysis of word order agreement and comparisons of the outputs between the baseline and the transfer learning system is made. The analysis and evaluation methods portray that neural machine translation using transfer learning improves the translation accuracy for Khasi, a low-resource language.

Keywords: Transfer learning; Neural machine translation; Khasi; Byte pair encoding; Low resource language

1. Introduction

Over the years, neural machine translations have improved the accuracy of various languages across the globe. The most prominent issue which affects the translation accuracy is the size of the corpora. This drastically affects neural machine translation for low-resource languages. Thus, getting an equivalent translation accuracy for low-resource languages compared to high-resource languages is challenging. However, applying recent techniques such as transfer learning to neural machine translation has shown promising results for many low-resource languages. The main pur-

* Corresponding author. Tel.: +0-000-000-0000 ; fax: +0-000-000-0000.
E-mail address: avhujon@gmail.com

10.1016/j.procs.2022.12.396

pose of our work is to improve the translation accuracy of neural machine translation of English to Khasi. The Khasi language is an Austroasiatic language with low resource constraints. It is classified under the North Mon-Khmer group of Austroasiatic languages. The Khasi alphabet was standardized in 1896 with 23 letters using the Latin Script. The word order of the Khasi language is Subject-Verb-Object(SVO). It is primarily spoken by the indigenous people of the state of Meghalaya, India. Focusing on low-resource languages of the northeastern region of India, substantial work has been reported for languages like Assamese, Manipuri, and some on the Mizo language. Among these works, Statistical machine translation[1], neural machine translation[2] and unsupervised neural machine translation[3] have been reported for Manipuri language. However, minimal work has been reported for machine translation of the Khasi language. An attempt at machine translation for the Khasi language has been explored[4], but due to the lack of significant size of the dataset, much improvement is still required to increase the translation accuracy. Many low-resource languages like Khasi face similar challenges due to low resource constraints in building substantial parallel corpora.

The existing works closely related to transfer learning machine translation and a few works on the machine translation of the English-Khasi language pair are discussed in Section 2. The framework and methods we based our experiments on to improve the accuracy and quality of the predicted outputs using neural machine translation with transfer learning concept for the language pair English-Khasi are discussed in Section 3. Next, Section 4 explains the datasets of our experiments and the method applied. The outputs of the baseline and the transfer learning models are analyzed and evaluated using human, statistical, and automatic evaluations in Section 5. The output texts of the two models are also analyzed based on word order agreement and comprehension. We also compare the two models and discuss the findings in detail. Section 6 finally concludes the paper.

2. Related Works

Machine translation using transfer learning has improved BLEU scores for various low-resource languages. One of the most significant reports is the transfer learning technique by Zoph et al.[5], which uses one parent model and applies the learned knowledge to various child models. The child models are initialized with the learned parameters from the parent model trained on the French-English dataset. The word embeddings of the target language(English) from the parent model are copied to the child model unmodified, while the child model's word embeddings of the source language are initialized by mapping to random weights of the source language(French) of the parent model. Considering the child model language pair is Hausa-English, then the word embeddings of Hausa are initialized to the random weights of French. The results of the child models for low-resource languages such as Hausa, Turkish, Uzbek, and Urdu, have shown significant improvement in the translation accuracy. The technique reported by Nguyen and Chiang[6] uses a similar method but applies the byte pair encoding technique to increase the overlap of related languages in the vocabulary. Transliteration is also applied, and the results show a more significant improvement in BLEU scores. The model uses global attention with two layers and 512-hidden-units. The results show that transfer learning is more effective when the languages of parent and child models are related to one another. A report by Kocmi and Bojar[7] shows that transfer learning can be made even more straightforward and can be applied to languages that do not share any relationship also. The parent and child models are implemented using a transformer model instead of the recurrent neural network model. The parent model is trained till it converge. The child model is trained from the parent model without resetting any training parameters for the rest of the training by switching the training corpus to the child language pair. An essential part of the training is the shared vocabulary. The vocabulary is generated by selecting an equal number of sentences for both the language pairs of the parent and child model. The results show improved BLEU scores even for languages not sharing any typical relationship. A study on the impact of using related languages in transfer learning is reported in [8]. Languages chosen for the study are Indian (Hindi, Marathi, Punjabi, Malayalam), Afro-Asiatic (Hausa, Somali), European(French, German, Luxembourgish), Slavic (Russian), Turkic (Turkish, Uzbek, Kazakh), and Austronesian (Indonesian, Javanese, Sundanese). Out of these, the parent language pairs chosen are Hindi, Turkish, Russian, German, and French. The child's languages are Luxembourgish, Hausa, Somali, Malayalam, and Punjabi. The target language for both parent and child models is English. An observation is reported that Hindi as a parent language improves the BLEU scores of the child models from +0.57 to +2.8 for all Indian languages compared to other parent languages. Even though French has the highest dataset compared to Hindi, it shows that size of the parent's dataset has a lesser impact than the relatedness of the languages. Machine translation for the Khasi language is still at its initial stage of progress. Statistical as well as neural machine translations have been experimented for the English-Khasi language pair[4] with significant translation accuracy. Cross-lingual language model have also been experimented for English-Khasi and Khasi-English translations[9] which achieved a translation accuracy of 39.63 and 32.69 BLEU scores respectively. Although there are a few reports of neural machine translation and statistical machine translation for English-Khasi[4, 9], transfer learning techniques have not been reported for machine translation of the Khasi language.

3. Research Methods

The notion of transfer learning is basically to use known knowledge to solve various problems in AI technology. Improving the learning of the target predictive function $F_T(.)$ in D_T is the main aim of transfer learning by using the knowledge gained in source D_S and T_S, where D_S is not the same as D_T, or T_S is not the same as T_T[10]. D_S, D_T are the domains of the source and target. T_S and T_T are the tasks of the source and target, respectively.

In consideration of the different conditions between the source and target tasks and domains, categorization of transfer learning can be done into three different categories[10]-inductive, transductive, and unsupervised. The domain may be the same, while the tasks for source and target are different in inductive transfer learning, whereas the tasks of the source and target are the same with different domains in transductive transfer learning. Unsupervised transfer learning is, however, very much similar to inductive transfer learning, where the target task is different from the source task, but they are related to each other in some way. The unsupervised transfer learning's main focus is on solving the unsupervised learning tasks of the target domain. Transfer learning applied in machine translation may be viewed as transductive transfer learning, where the task is machine translations for different domains such as different language pairs. The application of transfer learning in machine translations is often applied to improve the translation accuracy of low-resource languages with the help of high-resource languages. A standard method of applying transfer learning in machine translation is parameter initialization. The initialization approach first trains m source tasks $\{S_i\}_{i=1}^{m}$ of the neural networks and the learned parameters are used to initialize a neural network for a target task T. Then, the target neural network's parameters are updated using the labeled data which are available in T. There are generally two methods[11] to implement parameter initialization; the freezing method, in which some layers are frozen while training to preserve information, and fine-tuning method, where parameters are modified for training the child model. In some cases, both methods are also applicable, where some layers are frozen while new layers are fine-tuned. Translation of a low-resource language is challenging as increasing the size of the dataset requires a considerable amount of existing textual resources, which may not be available. With minimal dataset, a technique like transfer learning is a boon to neural machine translation for low-resource languages. Our method uses long short-term memory (LSTM) as the backbone architecture for neural machine translation. A layer of an LSTM architecture consists of LSTM cells. A unique feature of the LSTM cell is that each of these cells contains the input gate, forget gate, and output gate. These cells contribute to the performance of the LSTM network in remembering long sentences, which is lacking in a normal recurrent neural network. Matrix multiplication of the weights W^x with node values for the preceding layer, X^t is added with the matrix multiplication of the weights W^h and the given values of the hidden layer from the previous time step h^{t-1}. The activation function, g, is multiplied by the sum of the matrix multiplications. The result is then use as the input value for the cell as in equation 1 [12].

$$Input^t = g(W^x X^t + W^h h^{t-1}) \tag{1}$$

Matrices W^{xa}, W^{ha}, and W^{ma} is defined for each gate $a \in (input, forget, output)$ so as to compute the parameter value of the gate. This is done by multiplying the weights and node values in the previous layer x^t, the hidden layer h^{t-1} at the previous time step, and the memory states at the previous time step memory $t-1$,followed by an activation function h as in equation 2[12].

$$gate_a = h(W^{xa} X^t + W^{ha} h^{t-1} + W^{ma} memory^{t-1}) \tag{2}$$

The transfer learning technique applied is closely related to Kocmi and Bojar [7]. However, the vocabulary is created using subword byte pair encoding[13, 14] of the two pairs of languages. The language pair of the parent model is English-French and English-Khasi for the child model. The parent model's language pair is chosen since English, French, and Khasi share similar script (the Latin script) and word order (subject-verb-object). Since a joint vocabulary is required of the two pairs of languages of the parent and child models, the size of the vocabulary can be quite

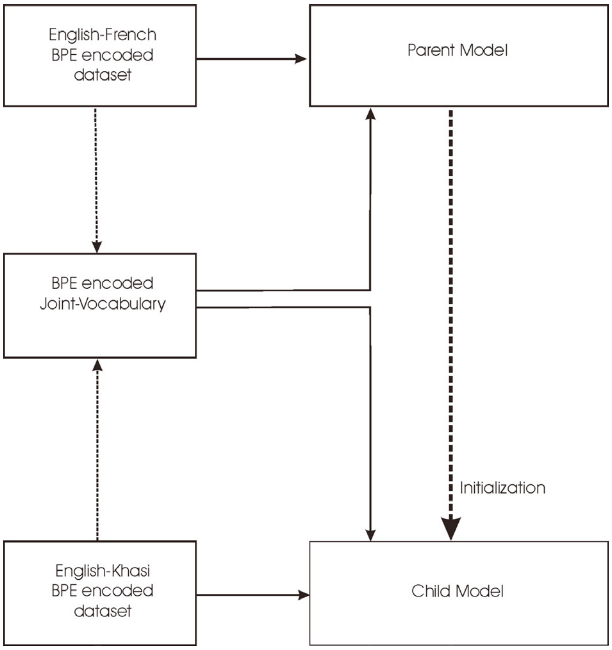

Fig. 1. NMT using Transfer Learning

large. To take care of out-of-vocabulary (OOV) words and to reduce the size of the vocabulary, subword units using subword byte pair encoding are applied. The parent and child datasets are first encoded using the subword byte pair encoding method. Instead of selecting an equal number of sentences from the parent and child datasets, the vocabulary is constructed using all the sentences in the train datasets of the languages of both models. We train the child model from the parent model. A few of the child model's parameters are fine-tuned by keeping most of the parameters, such as learning rate and dropout rate, the same as the parent model, whereas parameters such as the training steps are increased, and the dataset is switched to English-Khasi. The parent model is trained till convergence, followed by training of the child model while sharing the same byte pair encoded joint vocabulary. The proposed transfer learning approach for English to Khasi translation is shown in Fig. 1.

4. Experimental Setup

The baseline model is referred to as $NMT_{Baseline}$ and the neural machine translation using the transfer learning concept is referred to as NMT_{TL}. A comparative study is performed on experiments with the two models; the $NMT_{Baseline}$ and the NMT_{TL}. The baseline model uses a two-layer LSTM model with English-Khasi parallel corpora consisting of *41529* parallel sentences. The source of the English-Khasi parallel dataset are from the Bible[15, 16] and books that are digitized manually and aligned [17]. The same dataset is used for the baseline model as well as the transfer learning child model. The NMT_{TL} also uses the LSTM as a backbone architecture. The dataset of the parent model is extracted from the Europarl V7 French-English dataset [18]. The size distribution of both the datasets is given in Table 1.

Table 1. Datasets

Corpora	Train	Validation	Test
Baseline model	36601	1772	2656
Transfer learning Parent model	1000000	5000	500
Transfer learning Child model	36601	1772	2656

5. Results and Analysis

The performance of the two models are evaluated using automatic evaluation, statistical evaluation and human evaluation. The Bilingual Evaluation Understudy(BLEU) metric is used for automatic evaluation, and Precision, Recall and F-measure are used as the metrics for statistical evaluations by applying Precision as in equation 3, Recall as in equation 4 and F-Measure as in equation 5. The term *correct* is the number of words that are translated correctly in the translated text. *Output_length* is the number of words in the translated text, and *Reference_length* is the number of words in the reference text.

$$Precision = \frac{correct}{Output\_length} \tag{3}$$

$$Recall = \frac{correct}{Reference\_length} \tag{4}$$

$$F\text{-}Measure = \frac{Precision \times Recall}{(Precision + Recall)/2} \tag{5}$$

Human judgement or subjective evaluation is based on two important factors known as adequacy and fluency which are the two metrics applied for human evaluation. Adequacy is the amount of correct words translated in the target output and fluency is the correct ordering of words as per the grammatical rules.

The BLEU automatic score shows that NMT_{TL} is superior to $NMT_{Baseline}$ in performance, where the $NMT_{Baseline}$ gives a good score of 43.46, but the NMT_{TL} outperformed the $NMT_{Baseline}$ by 7.65 points with a score of 51.11. The scores are shown in Table 2. Statistical evaluation is performed on the output based on four categories of sen-

Table 2. Results of Automatic Evaluation

Model	BLEU Score	Brevity penalty	Ratio	Hypothesis Length	Reference Length
$NMT_{Baseline}$	43.46	0.953	0.956	47269	49557
NMT_{TL}	51.11	0.935	0.937	55180	58874

Table 3. Statistical Evaluation

		≤ 15			> 15 ≤ 25	
Model	Precision	Recall	F Measure	Precision	Recall	F Measure
$NMT_{Baseline}$	79.41%	69.05%	72.26%	52.38%	61.11%	56.41%
NMT_{TL}	94.44%	87.30%	88.89%	68.42%	72.22%	70.27%
		> 25 ≤ 50			> 50	
Model	Precision	Recall	F Measure	Precision	Recall	F Measure
$NMT_{Baseline}$	40.82%	46.51%	43.48%	61.54%	55.56%	58.39%
NMT_{TL}	51.11%	53.49%	52.27%	73.61%	73.61%	73.61%

tences grouped by the number of words in the input text. The scores are shown in Table 3. The statistical score for the $NMT_{Baseline}$ and NMT_{TL} shows that the translation quality is the highest among the four categories for short sentences of length, which are less than or equal to 15 words. However, the statistical score shows that both models performed similarly. The performances gradually decrease from short to longer sentences and slightly improve with long sentences of more than fifty words. However, the NMT_{TL} performed better than the $NMT_{Baseline}$ which achieved an F-measure score of 88.89% whereas the $NMT_{Baseline}$ achieved an F-measure score of 72.26%. The performance is lowest for sentences of length greater than twenty words and less than fifty words category with an F-measure of 52.27%. Subjective human evaluation is also performed on the output text based on the four categories of sentences similarly grouped as done for the statistical evaluation. In this evaluation which is shown in Table 4, it is found that

Table 4. Results of Human Evaluation based on group of Sentence Length in words

Model	≤ 15 adequacy	≤ 15 fluency	$> 15 \leq 25$ adequacy	$> 15 \leq 25$ fluency	$> 25 \leq 50$ adequacy	$> 25 \leq 50$ fluency	> 50 adequacy	> 50 fluency
$NMT_{Baseline}$	4.23	4.90	3.42	4.75	3.20	4.20	2.78	3.50
NMT_{TL}	4.85	5	3.66	4.75	3.98	4.50	3.68	4.30

the results of the evaluation show a correlation with a statistical evaluation where performances of short sentences are better than long sentences. Again, NMT_{TL} outperformed the $NMT_{Baseline}$ with adequacy of 4.85 and fluency of 5 for short sentences less than and equal to 15 words, which is the highest score for all four categories. With a closer analysis, it is found that the fluency of the output text gradually decreases from short to very long sentences, that is, $5, 4.75, 4.50$ and 4.30. The score is significantly high, which implies that the output predicted by the NMT_{TL} is fluent and comprehensibly understandable for all ranges of sentence length. Thus, all three methods of evaluation correlate with one another, where the NMT_{TL} achieves higher performance than the baseline model.

5.1. Sample Input-Output Analysis

Table 5. Sample Input-Output of Sentences having less than 15 number of words

Input	But he saved them, as he had promised
Reference	Hynrei u la pynim ïa ki, kumba u la kular
$NMT_{Baseline}$	Hynrei u la pynim ïa ki, kumba u la pynshitom
NMT_{TL}	Hynrei u la pyllait im ïa ki , kumba u la kular

Table 6. Sample Input-Output of Sentences more than 15 and less than 25 number of words

Input	On the Lord’s day the Spirit took control of me, and I heard a loud voice, that sounded like a trumpet, speaking behind me.
Reference	Ha ka sngi U Trai, U Mynsiem u la shim ïa nga, bad nga la ïohsngew ïa ka sur bajam, kaba sawa kum ka turoi, kaba kren na shadien jong nga.
$NMT_{Baseline}$	U Trai ka sngi U Trai u la wan ha nga, bad nga la ïohsngew ïa ka sur kaba jam, ba ka sawa kum ka turoi, ki kren ha hadien jong nga.
NMT_{TL}	Ha ka sngi U Trai U Mynsiem u la shim ïa nga , bad nga la ïohsngew ïa ka sur kaba jam , kata ka sawa kum ka turoi , kaba kren shadien jong nga .

Table 7. Sample Input-Output of Sentences more than 25 and less than 50 number of words

Input	The first one looked like a lion; the second looked like a bull; the third had a face like a human face; and the fourth looked like an eagle in flight.
Reference	Uba nyngkong u don ka dur kum u sing; uba ar u don ka dur kum u masi kyrtong; uba lai u don ka durkhmat kum ka durkhmat u briew; bad uba saw u don ka dur kum u pukni uba her.
$NMT_{Baseline}$	Uba nyngkong u syriem kum u sing uba ar iba la khmih kum u ’tiew lili. uba lai u la don ka durkhmat kum u traishnong, bad kaba saw ka la i kum u pukni kaba suh lyngkhuit.
NMT_{TL}	Uba nyngkong u la khmih kum u sing ; uba ar u la khmih kum u masi kyrtong ; uba lai u don ka durkhmat kum ka durkhmat briew ; bad kaba saw i kum u pukni ha ka lieng

Considering the sample output texts for short sentences in Table 5, the NMT_{TL} model translation is more accurate in terms of word order as well as comprehension. The word ‘promised’ in the input text is translated correctly, although

Table 8. Sample Input-Output of Sentences more 50 number of words

Input	Then I saw a Lamb standing in the centre of the throne, surrounded by the four living creatures and the elders. The Lamb appeared to have been killed. It had seven horns and seven eyes, which are the seven spirits of God that have been sent throughout the whole earth.
Reference	Nangta nga la ïohi ïa U Khunlangbrot uba ïeng hapdeng ka khet, ïa kaba la ker sawdong da ki saw tylli ki jingthaw ba-im bad ki nongsharai balang. U Khunlangbrot u la paw ba u la shah pynïap. U don hynñiew tylli ki reng bad hynñiew tylli ki khmat, kiba long ki hynñiew ki mynsiem U Blei ïa kiba la phah sha kylleng ka pyrthei.
$NMT_{Baseline}$	Nangta nga la ïohi shi kiaw ha ka pdeng jong ka khet, ba la ker sawdong da ki saw tylli ki jingthaw ba-im bad ki tymmen ki san. U Ittai u la paw ha ki hynñiew tylli ki reng bad hynñiew tylli ki khmat, kiba long ki hynñiew tylli ki mynsiem jong U Blei kiba la phah sha kylleng ka pyrthei baroh kawei.
NMT_{TL}	Nangta nga la ïohi ïa u Lamb uba ïeng ha ka dohnud jong ka khet , u kerkut da ki saw tylli ki jingthaw kiba im bad ki tymmen ki san . Ka Lamb ka la paw ban shah pynïap . Ka don hynñiew tylli ki reng bad hynñiew tylli ki khmat , kiba long ki hynñiew tylli ki mynsiem jong U Blei kiba la phah ha kylleng ka pyrthei.

the word ‘saved’ is translated to the phrase ‘pyllait im’, which is similar in meaning to the reference word ‘pynim’, while the $NMT_{Baseline}$ translated the word ‘promised’ to the word ‘pynshitom’ instead of the correct reference word ‘kular’. In the second sample output text in Table 6, the phrase ‘speaking behind me’ in the input text is translated more accurately by the NMT_{TL}. The fluency of the NMT_{TL} is slightly hindered by a few missing articles in the output text of the NMT_{TL}. The translation by the $NMT_{Baseline}$ model is inaccurate in many of the words in the input text. One such example is the word ‘Spirit’ which is translated to ‘Trai’ instead of the reference word ‘Mynsiem’. Similarly, in the phrase ‘On the Lord’s day the Spirit took control of me.’ in the input text where the correct phrase as in the reference text is ‘Ha ka sngi U Trai, U Mynsiem u la shim ia nga’, we find that this is one of the phrases on which the translation is completely accurate in terms of word order and comprehension by the NMT_{TL} while the $NMT_{Baseline}$ model fail to do so. The phrase ‘sounded like a trumpet’ is translated correctly and conforms to the word order by both the models. In the third sample sentence in Table 7, the first phrase of the input sentence, ‘The first one looked like a lion’, is translated correctly by the $NMT_{Baseline}$ as compared to NMT_{TL}. But the rest four phrases of the sentence are translated more correctly by the NMT_{TL} than $NMT_{Baseline}$. The NMT_{TL} also maintain the word order of the target language and achieve good comprehension. Referring to the output in Table 8, the word ‘Lamb’ appears to be an OOV word, the $NMT_{Baseline}$ model gives an incorrect translation as ‘kiaw’ and ‘ittai’ where the correct translated word as in the reference text is ‘khunlangbrot’, whereas NMT_{TL} translate this word with ‘Lamb’, the same English word itself as in the input text. The phrase ‘The Lamb appeared to have been killed’ is translated more correctly by NMT_{TL} as ‘Ka Lamb ka la paw ban shah pynıap’ while the $NMT_{Baseline}$ model fails to correctly translate this phrase. Overall, the quality of the output sentence of the NMT_{TL} is better than $NMT_{Baseline}$, which also conforms with the word order agreement of the target language and comprehension.

6. Conclusion

The analysis and evaluation performed on the output texts of the two models;$NMT_{Baseline}$ and NMT_{TL} have shown that the NMT_{TL} model is superior in performance compared to the $NMT_{Baseline}$ model. The results show that the accuracy of machine translation of low-resource languages can improve by implementing neural machine translation using the transfer learning technique. Subword byte pair encoding of both the datasets and vocabulary is one of the significant contributions to the results of our experiments. Utilizing a shared joint vocabulary is another factor that contributes to the translation accuracy achieved. The joint vocabulary, which consists of three languages, English, French, and Khasi, has contributed to the performance of the transfer learning model with a BLEU score of 51.11 compared to the baseline model with a 43.46 BLEU score. The translation accuracy of our models has shown a significant improvement compared to the existing BLEU score of 39.63[9] for the English-Khasi language pair. Results of both human and statistical evaluation of our models also agree with the automatic evaluation, which projects the transfer learning model NMT_{TL} using subword byte pair encoding as a better model and predicts more fluent sentences. Word order agreement analysis on the output texts also shows agreement to the word order of the target language, and the output sentences are comprehensibly understandable for variable sentence length. The results of our experiments indicate a satisfactory improvement in the translation accuracy of machine translation of the English-Khasi language pair.

References

[1] T. D. Singh, S. Bandyopadhyay, Manipuri-english bidirectional statistical machine translation systems using morphology and dependency relations, in: Proceedings of the 4th Workshop on Syntax and Structure in Statistical Translation, 2010, pp. 83–91.

[2] S. M. Singh, T. D. Singh, An empirical study of low-resource neural machine translation of manipuri in multilingual settings, Neural Computing and Applications 34. doi:10.1007/s00521-022-07337-8.

[3] S. M. Singh, T. D. Singh, Unsupervised neural machine translation for English and Manipuri, in: Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages, Association for Computational Linguistics, Suzhou, China, 2020, pp. 69–78.
URL https://aclanthology.org/2020.loresmt-1.10

[4] T. D. Singh, A. V. Hujon, Low resource and domain specific english to khasi smt and nmt systems, in: Proceedings of International Conference on Computational Performance Evaluation (ComPE-2020), 2020, pp. 733–737. doi:10.1109/ComPE49325.2020.9200059.

[5] B. Zoph, D. Yuret, J. May, K. Knight, Transfer learning for low-resource neural machine translation, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Austin, Texas, 2016, pp. 1568–1575. doi:10.18653/v1/D16-1163.
URL https://aclanthology.org/D16-1163

[6] T. Q. Nguyen, D. Chiang, Transfer learning across low-resource, related languages for neural machine translation, in: Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Asian Federation of Natural Language Processing, Taipei, Taiwan, 2017, pp. 296–301.
URL https://aclanthology.org/I17-2050

[7] T. Kocmi, O. Bojar, Trivial transfer learning for low-resource neural machine translation, CoRR abs/1809.00357. arXiv:1809.00357.
URL http://arxiv.org/abs/1809.00357

[8] R. Dabre, T. Nakagawa, H. Kazawa, An empirical study of language relatedness for transfer learning in neural machine translation, in: Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation, The National University (Phillippines), 2017, pp. 282–286.
URL https://aclanthology.org/Y17-1038

[9] N. D. J. Thabah, B. S. Purkayastha, Low resource neural machine translation from english to khasi: A transformer-based approach, in: Proceedings of the International Conference on Computing and Communication Systems: I3CS 2020, NEHU, Shillong, India, Vol. 170, Springer, 2021, p. 3.

[10] S. J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on Knowledge and Data Engineering 22 (2010) 1345–1359.

[11] Q. Yang, Y. Zhang, W. Dai, S. J. Pan, Transfer Learning in Natural Language Processing, Cambridge University Press, 2020, p. 234–256. doi:10.1017/9781139061773.020.

[12] P. Koehn, Neural Machine Translation, Cambridge University Press, Cambridge, 2020. doi:10.1017/9781108608480.

[13] R. Sennrich, B. Haddow, A. Birch, Neural machine translation of rare words with subword units, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Berlin, Germany, 2016, pp. 1715–1725. doi:10.18653/v1/P16-1162.
URL https://aclanthology.org/P16-1162

[14] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Computation 9 (8) (1997) 1735–1780. doi:10.1162/neco.1997.9.8.1735.
URL https://doi.org/10.1162/neco.1997.9.8.1735

[15] YouVersion, Khasiclbsi bible: Youversion.
URL https://www.bible.com/en-GB/bible/1865/GEN.1.KHASICLBSI

[16] YouVersion, Gnb bible: Youversion.
URL https://www.bible.com/en-GB/bible/296/GEN.1.GNB

[17] A. V. Hujon, T. D. Singh, Existing english to khasi translated documents for parallel corpora development: A survey, International Journal on Natural Language Computing (IJNLC) 7 (5) (2018) 81–91.

[18] P. Koehn, Europarl: A parallel corpus for statistical machine translation, in: Proceedings of Machine Translation Summit X: Papers, Phuket, Thailand, 2005, pp. 79–86.
URL https://aclanthology.org/2005.mtsummit-papers.11