

# GX@DravidianLangTech-EACL2021: Multilingual Neural Machine Translation and Back-Translation

Wanying Xie

Beijing Language and Culture University, China

xiewanying07@gmail.com

## Abstract

In this paper, we describe the GX system in the EACL2021 shared task on machine translation in Dravidian languages. Given the low amount of parallel training data, We adopt two methods to improve the overall performance: (1) *multilingual translation*, we use a shared encoder-decoder multilingual translation model handling multiple languages simultaneously to facilitate the translation performance of these languages; (2) *back-translation*, we collected other open-source parallel and monolingual data and apply back-translation to benefit from the monolingual data. The experimental results show that we can achieve good translation results in these Dravidian languages and rank first in the four translation directions on the ranklist.

## 1 Introduction

In recent years, encoder-decoder based on neural machine translation (NMT) has become the mainstream paradigm in the field of machine translation (Kalchbrenner and Blunsom, 2013; Sutskever et al., 2014; Bahdanau et al., 2015; Gehring et al., 2017). The Transformer model (Vaswani et al., 2017), which is completely based on self-attention mechanism, is the best model architecture for translation performance. However, these data-driven neural machine translation depends heavily on large-scale annotated parallel data, the translation performance of low-resource languages remains a concern (Zoph et al., 2016). One such language family is the Dravidian family of languages, which are under-resourced in speech and natural language processing (Chakravarthi et al., 2021). The Dravidian languages are first attested in the 6th century BCE as Tamil<sup>1</sup> script inscribed on the cave walls and pottery in the Madurai and Tirunelveli districts of Tamil Nadu. The Dravidian languages with the

<sup>1</sup>also called Damili or Tamil Bhrami

most speakers are Tamil, Kannada and Malayalam, of which Tamil have long literary traditions from 600 BCE.

In this paper, we adopt two approaches to get a reliable translation for improving access to and production of information for monolingual speakers of Dravidian languages. The first approach is multilingual translation. Compared with translating from a resource-poor language to English, translating from English to a resource-poor language is much more difficult. Translating from a resource-poor language to English, can apply transfer learning to make use of large parallel corpora between English and other languages (Zoph et al., 2016; Chakravarthi, 2020). However, the situation of translating from English to a resource-poor language is tough because there is little parallel data for the target language and other languages.

In the case of low resources, the multilingual translation method is often better than the performance of individual language pair, because too little data makes it difficult for the model to converge to the best (Dong et al., 2015; Tan et al., 2019; Chakravarthi et al., 2019b,a). The other approach is back-translation. Data augmentation methods can help alleviate this resource shortage by increasing the amount of data. Back-translation is a typical method of data augmentation that can enrich training data with monolingual data. These work are often based on automatically creating pseudo-parallel sentences through back-translation (Irvine and Callison-Burch, 2013; Sennrich et al., 2016a; Feldman and Coto-Solano, 2020).

For the EACL2021 shared task on machine translation in Dravidian languages, we present our results on the four language pairs: English-Tamil, English-Malayalam, English-Telugu, and Tamil-Telugu. Overall, the four language pairs can be divided into two subtasks: *between English and Dravidian languages* and *between Dravidian Lan-*

guages. To get better performance, we implement a translation system for each subtask. Since the systems are not constrained using exclusive data provided by the task’s organization, we collect other open-source monolingual data of Tamil, Malayalam, and Telugu. We use back-translation to create pseudo-parallel sentences. Based on official and pseudo-parallel data, we train multilingual models to perform the translation task.

## 2 Background

We briefly describe neural machine translation model here. Given the source sentence  $\mathbf{x} = (x_1, \dots, x_I)$  and the corresponding target sentence  $\mathbf{y} = (y_1, \dots, y_J)$ , a standard NMT model directly optimizes the conditional probability:

$$P(\mathbf{y} | \mathbf{x}; \theta) = \prod_{j=1}^J P(y_j | \mathbf{y}_{<j}, \mathbf{x}; \theta) \quad (1)$$

where  $\theta$  is the set of the model parameters and  $\mathbf{y}_{<j}$  denotes the partial translation. The architecture of machine translation model is the neural network based encoder-decoder framework (Sutskever et al., 2014). The input sentence  $\mathbf{x}$  will be first converted to a sequence of vectors and fed into the encoder. Then the Encoder generates a context vector which is an accumulation of the hidden state information related to each source language word. The decoder decodes the target language words from this context vector. Typically, this framework can be implemented as recurrent neural network (RNN) (Bahdanau et al., 2015), convolutional neural network (CNN) (Gehring et al., 2017) and Transformer (Vaswani et al., 2017). The parameters of the NMT model are trained to maximize the likelihood of a set of training examples  $L = \{[\mathbf{x}^m, \mathbf{y}^m]\}_{m=1}^M$ :

$$\mathcal{L}(\theta) = \arg \max_{\theta} \sum_{m=1}^M \log P(\mathbf{y}^m | \mathbf{x}^m; \theta) \quad (2)$$

where  $M$  is the amount of training samples.

## 3 Data Preparation

The data of Dravidian languages provided by the organizers are scarce, with only tens or hundreds of thousands of parallel sentences per language pair, which is shown in Table 1. In general, it is difficult to train a well-performing translation model with such a small amount of data, because it

Official	En-Ta	En-Te	En-MI	Ta-Te
Release	28417	23222	382868	17155

Table 1: Statistics of the official published data for each language pair.

Parallel	En-Ta	En-Te	En-MI
Wiki Titles v2	102142	-	-
PMIndia v1	33669	40283	33669
Tanzil v1	93540	-	187081
NLPC_UOM	8949	-	-
UFAL EnTam	166871	-	-
Monolingual	Ta	Te	MI
PMIndia v1	97216	118274	88244

Table 2: Statistics of other open source datasets, including parallel and monolingual data.

is difficult for the model to learn enough translation knowledge and ability.

In this work, all proposed systems are allowed to use data other than those provided by the task organizer. Therefore, We use other open-source datasets, including Wiki Titles v2<sup>2</sup>, PMIndia v1<sup>3</sup>, Tanzil v1 (Tiedemann, 2012), The NLPC.UOM En-Ta corpus and glossary (v1.0.3) (Fernando et al., 2020) and The UFAL EnTam corpus (Ramasamy et al., 2012). As monolingual data, we use the data of PMIndia v1. See Table 2 for statistics information.

Sentences of English were tokenized by the Moses scripts<sup>4</sup>. All non-English data were tokenized by the Indic NLP Library tool (Kunchukuttan, 2020), which is a library to provide a general solution to very commonly required toolsets for Indian language text. Then, all sentences further segmented into subword symbols using Byte-Pair Encoding (BPE) rules (Sennrich et al., 2016b) with 40K merge operations for all languages jointly.

## 4 System Description

In this section, we will describe our translation system with the adapted methods in detail for different categories of translation direction, including *between English and Dravidian languages* and *between Dravidian Languages*.

<sup>2</sup><http://data.statmt.org/wikititles/v2/>

<sup>3</sup><http://data.statmt.org/pmindia/>

<sup>4</sup><http://www.statmt.org/moses/>

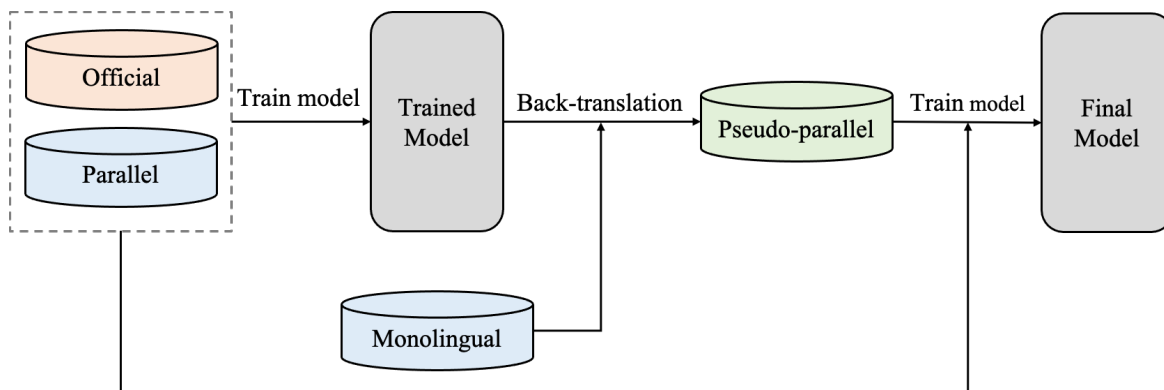


Figure 1: The whole training process of the proposed system. The orange data represents official data shown in Table 1, which is released by the organizers. The blue datasets represent other open-source data shown in Table 2. The green data represents pseudo-parallel data which is generated by applying back-translation on monolingual data.

#### 4.1 Between English and Dravidian Languages

In this part, there are three language pairs: English-Tamil, English-Malayalam, and English-Telugu. There is a one-to-three multilingual neural machine translation model for these three language pairs.

As shown in Figure 1, we first train a Three-to-En multilingual neural machine translation model by the sum of the data of the organizer and we collected (that is, the sum data of Official data in Table 1 and Parallel data in Table 2). With this well-trained translation model, we adapt back-translation method on the monolingual data of Tamil, Malayalam, and Telugu (that is, Monolingual data in Table 2) to generate pseudo-parallel sentences. Finally, we train an English-to-Three multilingual translation model based on the sum of the real and pseudo parallel corpora as our final model.

#### 4.2 Between Dravidian Languages

Tamil-Telugu is the most difficult one because this language pair has very little parallel and monolingual data. There is a four-to-four multilingual neural machine translation model for this language pair.

It has been shown that it is helpful to improve the overall performance when similar languages or languages in the same language family are trained together (Tan et al., 2019). Therefore, to facilitate this low resource translation language pair, we combine all the bilingual corpora to train a multilingual model. As shown in Figure 1, first, on the parallel data released by the organizer and we collected (that is, the sum

data of Official data in Table 1 and Parallel data in Table 2), we train a four-to-four multilingual neural machine translation model with eight directions of English $\leftrightarrow$ Tamil, English $\leftrightarrow$ Malayalam, English $\leftrightarrow$ Telugu and Tamil $\leftrightarrow$ Telugu. Then, pseudo-parallel sentences were generated by this multilingual on monolingual data of Telugu. Finally, we train a multilingual translation model of these eight directions based on the sum of all languages data as our final model.

## 5 Experiments

We first introduce the training details and empirically evaluate the systems in two scenarios.

### 5.1 Training Details

We implemented all proposed systems on the advanced Transformer (Vaswani et al., 2017) model using the open-source toolkit *Fairseq-py* (Ott et al., 2019), which consists of 6 stacked encoder/decoder layers with the layer size being 512 and 4 heads in each attention layer and feedforward network size being 1024. All the models were trained on 4 GeForce RTX 3090, with 24GB memory, where each was allocated with a batch size of 2,000 tokens for these two scenarios. We trained the model with dropout = 0.3 and using Adam optimizer (Kingma and Ba, 2015) with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.98$ , and  $\epsilon = 10^{-9}$ . For evaluation, we average the last 5 checkpoints saved and use beam search with a beam size of 5 and length penalty  $\alpha = 0.6$ . The translation quality of development sets was evaluated using the multi-bleu.pl script (Papineni et al., 2002) based on n-gram matching with n up to 4.

<b>Systems</b>	En-Ta	$\Delta$	En-Te	$\Delta$	En-MI	$\Delta$	Ta-Te	$\Delta$
Individual	32.92	-	34.62	-	24.82	-	28.63	-
Multilingual	47.97	15.05	40.39	5.77	25.01	0.19	54.97	26.34
+Monolingual	50.51	2.53	45.35	4.96	25.85	0.84	58.23	3.26
+Checkpoint Avg	50.70	0.21	45.65	0.30	25.87	0.02	58.27	0.04

Table 3: Results of development sets measured in BLEU on the Dravidian languages Translation task.  $\Delta$  means the difference between the current result and the previous row. Individual means an separate NMT model for each language pair without any other approaches. Multilingual means Multi-NMT models for the two tasks without other approaches.

## 5.2 Systems

The specific meanings of each system in the experiment are as follows:

**Individual** A NMT model is trained for each language pair. Therefore, there are 4 different models for the translation directions.

**Multilingual** (Johnson et al., 2017) Handling multiple languages in a single transformer model which contains one encoder and one decoder with a special language indicator added to the input sentence.

**+Monolingual** Add the monolingual data on the Multilingual system. It is the back-translation method, which are introduced in Section 4.

**+Checkpoint Avg** On the basis of +Monolingual system, we average the last 5 checkpoints to get the final model.

## 5.3 Results

In this section we will introduce the results of development sets and test sets.

**Development Sets** The development sets results of the four directions are reported in Table 3. We will discuss the two parts of the results separately.

Between English and Dravidian Languages, there is a huge boost between multilingual and individual, which up to 15.05 in En→Ta. This also proves that it is difficult to achieve good results that train an individual translation model only on low-resource language pairs. While multilingual neural machine translation, especially training on multiple similar languages, will greatly improve the translation performance of low-resource language pairs. Moreover, the addition of pseudo-parallel corpus generated by monolingual data can also bring improvement, but the improvement is relatively weak. This may be because the monolingual data is not enough, and the parallel sentence requirements of training excellent models still cannot be met. Besides, checkpoint average also improves the trans-

<b>Systems</b>	En-Ta	En-Te	En-MI	Ta-Te
Ours	36.66	38.86	19.84	35.29

Table 4: Results of test sets measured in BLEU on the Dravidian languages translation task.

lation performance but the increase is very low. It is worth noting that the En→MI is not highly promoted by our method. This may be because the parallel sentences are more than other languages so that it is difficult to gain knowledge from multilingual translation and other approaches to improve.

Between Dravidian Languages, the translation performance of Ta→Te benefit a lot from multilingual translation. That is because individual baseline trained on extremely low resources hard to get a good performance and the directions training samples of Ta→En and En→Te may boost the performance of Ta→Te when training a multilingual neural machine translation model. At the same time, other approaches such as back-translation and checkpoint average also improve the overall translation quality.

**Test Sets** We evaluate the test set using the model that works best on the validation set, the final model that +Checkpoint Avg produced. The results are shown in Table 4. Surprisingly, all systems score low on the test sets, presumably because the validation set was easy and the test set was hard. In the ranklist given by the organizer, our system ranks first in all four translation directions, but our system did not perform particularly well. Other competitors’ systems also struggled to perform well, which shows that Dravidian languages translation is a difficult task. To obtain a satisfactory translation result, we still have a lot of effort to do.

## 6 Related Work

There are several ways focusing on low-resource translation, including: (a) learning knowledge from high-resource pairs to promotes low-resource translation, such as transfer learning, transferring proper parameters (Zoph et al., 2016; Gu et al., 2018b), lexical knowledge (Nguyen and Chiang, 2017; Lakew et al., 2018) and syntactic knowledge (Gu et al., 2018a; Murthy et al., 2019) from high-resource language, and multilingual translation, including sharing encoders (Dong et al., 2015), sharing decoders (Zoph et al., 2016), sharing sublayers (Firat et al., 2016) and an universal Multi-NMT model with a target language token to indicate the translation direction (Ha et al., 2016; Johnson et al., 2017); (b) data augmentation in NMT, such as back-translation (BT) translates target-language monolingual text to create synthetic parallel sentences (Irvine and Callison-Burch, 2013; Sennrich et al., 2016a) and adding an auxiliary auto-encoding task on monolingual data (Cheng et al., 2016; Currey et al., 2017).

Also, Indian languages have been studied in recent years (Tennage et al., 2017; Chakravarthi et al., 2018; Kumar and Singh, 2019; Barrault et al., 2020). Choudhary et al. (2018) show that subword units method is helpful when translation Tamil language. Escolano et al. (2020) adopt a multilingual adaptation method to benefit from the positive transfer from high resource languages in English and Tamil translation. Bao et al. (2020) explore pre-training and iterative back translation for low-resource English-Tamil. Chen et al. (2020) explore self-supervised model pretraining, multilingual models, data augmentation, and reranking techniques to solve the low resource problem.

## 7 Conclusion

In this paper, we describe the GX system in the Dravidian translation task of EACL2021. We use two main methods, multilingual translation and back-translation, to enable the model to benefit from multiple languages joint learning and monolingual data. Experimental results show that the system can successfully introduce multilingual neural machine translation and back-translation on monolingual data and improve the performance of the system.

## Acknowledgments

We would like to thank the anonymous reviewers for their insightful comments. We thank Shuhao

Gu for his helpful suggestions and inspiration.

## References

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. [Neural machine translation by jointly learning to align and translate](#). In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Calvin Bao, Yow-Ting Shiue, Chujun Song, Jie Li, and Marine Carpuat. 2020. [The university of maryland’s submissions to the WMT20 chat translation task: Searching for more data to adapt discourse-aware neural machine translation](#). In *Proceedings of the Fifth Conference on Machine Translation, WMT@EMNLP 2020, Online, November 19-20, 2020*, pages 456–461. Association for Computational Linguistics.
- Loïc Barrault, Magdalena Biesialska, Ondrej Bojar, Marta R. Costa-jussà, Christian Federmann, Yvette Graham, Roman Grundkiewicz, Barry Haddow, Matthias Huck, Eric Joanis, Tom Kocmi, Philipp Koehn, Chi-kiu Lo, Nikola Ljubesic, Christof Monz, Makoto Morishita, Masaaki Nagata, Toshiaki Nakazawa, Santanu Pal, Matt Post, and Marcos Zampieri. 2020. [Findings of the 2020 conference on machine translation \(WMT20\)](#). In *Proceedings of the Fifth Conference on Machine Translation, WMT@EMNLP 2020, Online, November 19-20, 2020*, pages 1–55. Association for Computational Linguistics.
- Bharathi Raja Chakravarthi. 2020. *Leveraging orthographic information to improve machine translation of under-resourced languages*. Ph.D. thesis, NUI Galway.
- Bharathi Raja Chakravarthi, Mihael Arcan, and John P. McCrae. 2018. [Improving wordnets for under-resourced languages using machine translation](#). In *Proceedings of the 9th Global Wordnet Conference*, pages 77–86, Nanyang Technological University (NTU), Singapore. Global Wordnet Association.
- Bharathi Raja Chakravarthi, Mihael Arcan, and John P. McCrae. 2019a. [WordNet gloss translation for under-resourced languages using multilingual neural machine translation](#). In *Proceedings of the Second Workshop on Multilingualism at the Intersection of Knowledge Bases and Machine Translation*, pages 1–7, Dublin, Ireland. European Association for Machine Translation.
- Bharathi Raja Chakravarthi, Ruba Priyadarshini, Shubhanker Banerjee, Richard Saldhana, John Philip McCrae, Anand Kumar M, Parameswari Krishnamurthy, and Melvin Johnson. 2021. [Findings of the shared task on Machine Translation in Dravidian languages](#). In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

- Bharathi Raja Chakravarthi, Ruba Priyadharshini, Bernardo Stearns, Arun Jayapal, Sridevy S, Michael Arcan, Manel Zarrouk, and John P McCrae. 2019b. [Multilingual multimodal machine translation for Dravidian languages utilizing phonetic transcription](#). In *Proceedings of the 2nd Workshop on Technologies for MT of Low Resource Languages*, pages 56–63, Dublin, Ireland. European Association for Machine Translation.
- Peng-Jen Chen, Ann Lee, Changhan Wang, Naman Goyal, Angela Fan, Mary Williamson, and Jiatao Gu. 2020. [Facebook ai’s WMT20 news translation task submission](#). In *Proceedings of the Fifth Conference on Machine Translation, WMT@EMNLP 2020, Online, November 19-20, 2020*, pages 113–125. Association for Computational Linguistics.
- Yong Cheng, Wei Xu, Zhongjun He, Wei He, Hua Wu, Maosong Sun, and Yang Liu. 2016. [Semi-supervised learning for neural machine translation](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics.
- Himanshu Choudhary, Aditya Kumar Pathak, Rajiv Ratan Saha, and Ponnurangam Kumaraguru. 2018. [Neural machine translation for english-tamil](#). In *Proceedings of the Third Conference on Machine Translation: Shared Task Papers, WMT 2018, Belgium, Brussels, October 31 - November 1, 2018*, pages 770–775. Association for Computational Linguistics.
- Anna Currey, Antonio Valerio Miceli Barone, and Kenneth Heafield. 2017. [Copied monolingual data improves low-resource neural machine translation](#). In *Proceedings of the Second Conference on Machine Translation, WMT 2017, Copenhagen, Denmark, September 7-8, 2017*, pages 148–156. Association for Computational Linguistics.
- Daxiang Dong, Hua Wu, Wei He, Dianhai Yu, and Haifeng Wang. 2015. [Multi-task learning for multiple language translation](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*, pages 1723–1732.
- Carlos Escolano, Marta R. Costa-jussà, and José A. R. Fonollosa. 2020. [The TALP-UPC system description for WMT20 news translation task: Multilingual adaptation for low resource MT](#). In *Proceedings of the Fifth Conference on Machine Translation, WMT@EMNLP 2020, Online, November 19-20, 2020*, pages 134–138. Association for Computational Linguistics.
- Isaac Feldman and Rolando Coto-Solano. 2020. [Neural machine translation models with back-translation for the extremely low-resource indigenous language bribri](#). In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 3965–3976. International Committee on Computational Linguistics.
- Aloka Fernando, Surangika Ranathunga, and Gihan Dias. 2020. [Data augmentation and terminology integration for domain-specific sinhala-english-tamil statistical machine translation](#). *CoRR*, abs/2011.02821.
- Orhan Firat, Kyunghyun Cho, and Yoshua Bengio. 2016. [Multi-way, multilingual neural machine translation with a shared attention mechanism](#). In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 866–875.
- Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N. Dauphin. 2017. [Convolutional sequence to sequence learning](#). In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, pages 1243–1252.
- Jiatao Gu, Hany Hassan, Jacob Devlin, and Victor O. K. Li. 2018a. [Universal neural machine translation for extremely low resource languages](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pages 344–354.
- Jiatao Gu, Yong Wang, Yun Chen, Victor O. K. Li, and Kyunghyun Cho. 2018b. [Meta-learning for low-resource neural machine translation](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 3622–3631. Association for Computational Linguistics.
- Thanh-Le Ha, Jan Niehues, and Alexander H. Waibel. 2016. [Toward multilingual neural machine translation with universal encoder and decoder](#). *CoRR*, abs/1611.04798.
- Ann Irvine and Chris Callison-Burch. 2013. [Combining bilingual and comparable corpora for low resource machine translation](#). In *Proceedings of the Eighth Workshop on Statistical Machine Translation, WMT@ACL 2013, August 8-9, 2013, Sofia, Bulgaria*, pages 262–270. The Association for Computer Linguistics.
- Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda B. Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2017. [Google’s multilingual neural machine translation system: Enabling zero-shot translation](#). *TACL*, 5:339–351.

- Nal Kalchbrenner and Phil Blunsom. 2013. [Recurrent continuous translation models](#). In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1700–1709.
- Diederik P. Kingma and Jimmy Ba. 2015. [Adam: A method for stochastic optimization](#). In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Amit Kumar and Anil Kumar Singh. 2019. [NLPRL at WAT2019: transformer-based tamil - english indic task neural machine translation system](#). In *Proceedings of the 6th Workshop on Asian Translation, WAT@EMNLP-IJCNLP 2019, Hong Kong, China, November 4, 2019*, pages 171–174. Association for Computational Linguistics.
- Anoop Kunchukuttan. 2020. The IndicNLP Library. [https://github.com/anoopkunchukuttan/indic\\_nlp\\_library/blob/master/docs/indicnlp.pdf](https://github.com/anoopkunchukuttan/indic_nlp_library/blob/master/docs/indicnlp.pdf).
- Surafel Melaku Lakew, Aliia Erofeeva, Matteo Negri, Marcello Federico, and Marco Turchi. 2018. [Transfer learning in multilingual neural machine translation with dynamic vocabulary](#). *CoRR*, abs/1811.01137.
- V. Rudra Murthy, Anoop Kunchukuttan, and Pushpak Bhattacharyya. 2019. [Addressing word-order divergence in multilingual neural machine translation for extremely low resource languages](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 3868–3873. Association for Computational Linguistics.
- Toan Q. Nguyen and David Chiang. 2017. [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)*, pages 296–301, Taipei, Taiwan. Asian Federation of Natural Language Processing.
- Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. 2019. [fairseq: A fast, extensible toolkit for sequence modeling](#). In *Proceedings of NAACL-HLT 2019: Demonstrations*.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: a method for automatic evaluation of machine translation](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia, PA, USA*, pages 311–318.
- Loganathan Ramasamy, Ondřej Bojar, and Zdeněk Žabokrtský. 2012. [Morphological processing for english-tamil statistical machine translation](#). In *Proceedings of the Workshop on Machine Translation and Parsing in Indian Languages (MTPIL-2012)*, pages 113–122.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016a. [Improving neural machine translation models with monolingual data](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016b. [Neural machine translation of rare words with subword units](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*.
- Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. [Sequence to sequence learning with neural networks](#). In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 3104–3112.
- Xu Tan, Jiale Chen, Di He, Yingce Xia, Tao Qin, and Tie-Yan Liu. 2019. [Multilingual neural machine translation with language clustering](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 963–973.
- Pasindu Tennage, Prabath Sandaruwan, Malith Thilakarathne, Achini Herath, Surangika Ranathunga, Sanath Jayasena, and Gihan Dias. 2017. [Neural machine translation for sinhala and tamil languages](#). In *2017 International Conference on Asian Language Processing, IALP 2017, Singapore, December 5-7, 2017*, pages 189–192. IEEE.
- Jörg Tiedemann. 2012. [Parallel data, tools and interfaces in OPUS](#). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey, May 23-25, 2012*, pages 2214–2218. European Language Resources Association (ELRA).
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 5998–6008.
- Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 2016. [Transfer learning for low-resource neural machine translation](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural*

*Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 1568–1575. The Association for Computational Linguistics.