

# The ADAPT System Description for the WMT20 News Translation Task

Venkatesh Balavadhani Parthasarathy<sup>‡</sup>, Akshai Ramesh<sup>‡</sup>, Rejwanul Haque and Andy Way

The ADAPT Centre, <sup>‡</sup>School of Computing  
Dublin City University, Dublin, Ireland

venkatesh.balavadhaniparthasa2, akshai.ramesh2@mail.dcu.ie  
rejwanul.haque, andy.way@adaptcentre.ie

## Abstract

This paper describes the ADAPT Centre’s submissions to the WMT20 News translation shared task for English-to-Tamil and Tamil-to-English. We present our machine translation (MT) systems that were built using the state-of-the-art neural MT (NMT) model, Transformer. We applied various strategies in order to improve our baseline MT systems, e.g. monolingual sentence selection for creating synthetic training data, mining monolingual sentences for adapting our MT systems to the task, hyperparameters search for Transformer in low-resource scenarios. Our experiments show that adding the aforementioned techniques to the baseline yields an excellent performance in the English-to-Tamil and Tamil-to-English translation tasks.

## 1 Introduction

The ADAPT Centre participated in the News Translation Shared Task of the Fifth Conference of Machine Translation (WMT20) in the English-to-Tamil and Tamil-to-English language directions. To build our neural MT systems we used the Transformer model (Vaswani et al., 2017). Our strategies to build the competitive MT systems for the task include applying the state-of-the-art data augmentation approaches (e.g. (Sennrich et al., 2016a; Caswell et al., 2019)), selecting “pseudo” in-domain monolingual sentences for the creation of synthetic bitexts, mining monolingual source and target sentences for the adaptation of neural MT (NMT) systems, finding the optimal set of hyperparameters for Transformer as far as low-resource translation is concerned.

The remainder of the paper is organized as follows. In Section 2, we present our approaches for the MT system building. Section 3 first presents details of the data sets used and then presents the evaluation results with some discussions, while

Section 4 concludes our work with avenues for future work.

## 2 Our Strategies to improve MT Systems

### 2.1 Data Augmentation

The data augmentation methods (Sennrich et al., 2016a; Zhang and Zong, 2016; Burlot and Yvon, 2018; Poncelas et al., 2018; Bogoychev and Sennrich, 2019; Caswell et al., 2019; Chen et al., 2019), which usually employ the unlabeled monolingual data in addition to limited bitexts, can positively impact the MT system’s performance and are very popular among the MT developers and researchers (Barrault et al., 2019). In other words, use of augmented bitexts that include synthetic data to improve a NMT system is nowadays a common practice, especially in the under-resource scenarios. The synthetic training data whose target-side sentences are original is more effective for domain text translation and generation of fluent translations. In this task, in order to improve our baseline Transformer models, we augmented our training data with the target-original synthetic data. As in Caswell et al. (2019), in order to let the NMT model know that the given source is synthetic, we tag the back-translated source sentences with an extra token.

Note that we also tried applying the so-called self-training<sup>1</sup> strategy (Ueffing et al., 2007) to improve our NMT systems. However, this method does not bring any improvements in the Tamil-to-English translation task, and deteriorates the performance of the MT systems in the English-to-Tamil translation task.

Iterative generation and training on synthetic data can yield increasingly better NMT systems,

<sup>1</sup>Synthetic data for training is created by the MT system itself (i.e. source-side is original) (Zhang and Zong, 2016; Burlot and Yvon, 2018).

especially in low-resource scenarios (Hoang et al., 2018; Chen et al., 2019). Similarly, in order to produce our final English-to-Tamil and Tamil-to-English MT systems, we performed iterative training by back-translating new monolingual data with the updated MT system and appending the resultant synthetic data to the original training data in each iteration.

## 2.2 Selecting pseudo In-Domain Sentences

In an attempt to improve the quality of our NMT engines, we extracted monolingual sentences from large monolingual data that are similar to the styles of the in-domain data. Sentences of a large monolingual corpus similar to the in-domain sentences when selected based on the perplexity according to an in-domain language model were found to be effective in MT (Gao et al., 2002; Yasuda et al., 2008; Foster et al., 2010; Axelrod et al., 2011; Toral, 2013). As for NMT training, we believe that synthetic parallel data created using pseudo in-domain sentences can be better alternatives than those selected randomly. Accordingly, we select “pseudo” in-domain sentences from a large monolingual corpus based on the perplexity scores according to the in-domain language models. The extracted sentences are then back-translated with a target-to-source MT system to form synthetic training data.

## 2.3 Mining Monolingual Sentences for the Adaptation of the NMT models

Chinea-Ríos et al. (2017) demonstrated that in case of specialised domains or low-resource scenarios where parallel corpora are scarce sentences of a large monolingual data that are more related to the test set sentences to be translated could be effective for fine-tuning the original general domain NMT model. They select those instances from large monolingual corpus whose vector-space representation is similar to the representation of the test set instances. The selected sentences are then automatically translated by an NMT system that is trained on a general domain data. Finally, the NMT system is fine-tuned with the resultant synthetic data. In a similar line of research, it has also been shown that an NMT system built on general domain data can be fine-tuned using just a few sentences (Farajian et al., 2017, 2018; Wuebker et al., 2018; Huck et al., 2019).

In our case, since English–Tamil is a low-resource language-pair and have a little amount

of bitexts pertaining to the targeted domain (News), we followed Chinea-Ríos et al. (2017) and mined those sentences from large monolingual data that can be beneficial for fine-tuning the original NMT models. In addition to mining source-side sentences (Chinea-Ríos et al., 2017), we also mined target language sentences from large monolingual corpus (Huck et al., 2019) when English is the source language. However, our selection methods are different to those of the other papers (Chinea-Ríos et al., 2017; Farajian et al., 2017, 2018; Wuebker et al., 2018; Huck et al., 2019) and are described below.

Terms are usually indicators of the nature of a domain and play a critical role in domain-specific MT (Haque et al., 2020). The target translation could lose its meaning if the terminology translation is not dealt with care. Therefore, we focused on mining those sentences from a large monolingual corpus that contain domain terms. For this, we made use the approach of Rayson and Garside (2000); Haque et al. (2014, 2018) for identifying terms in the test set which is to be translated. This term extraction method performs well even on a small amount of sentences (Haque et al., 2014, 2018). The goal is to identify those words which are most indicative (or characteristic) of the test corpus compared to a reference corpus. Haque et al. (2014, 2018) used a large corpus which is generic in nature as a reference corpus. We adopted their approach and used a large generic corpus in order to identify terms in the test set. Additionally, in our second setup, we used the training set on which the NMT systems were trained as the reference corpus. The intuition is to extract those terms or sequence of words from the test set that do not occur or rarely occur in the training set and convey representativeness of the test set. We merged the two sets of terms extracted following the two setups above. Given the resultant list of terms, we mine sentences from monolingual corpus.

We observed that the WMT20 News development text contains many named entities (NEs) and many of them are out-of-vocabulary items. We also found that our initial MT systems miserably failed to translate many NEs. Therefore, we used Stanford named entity recogniser (NER)<sup>2</sup> (Finkel et al., 2005) in order to identify NEs in the English test set. As above, we used the extracted NEs in order to mine sentences from a large monolingual corpus.

<sup>2</sup><https://nlp.stanford.edu/software/CRF-NER.html>

We build an English-to-Tamil transliteration system and the extracted English NEs were transliterated into Tamil. Note that we took 5-best Tamil translations for an English NE as in [Huck et al. \(2019\)](#). These Tamil NEs were then used to mine Tamil sentences from a large target monolingual corpus.

In order to build the English-to-Tamil transliteration system, we used the 2016 Named Entity Transliteration Shared Task (NEWS) dataset<sup>3</sup> ([Duan et al., 2016](#)). We used our in-house machine transliteration tool ([Haque et al., 2009](#)) in order to prepare the English-to-Tamil transliteration system.

We could not apply this strategy in the Tamil-to-English translation task since there is no publicly available NER for Tamil. The source and target sentences that have been mined are translated with the final source-to-target and target-to-source NMT systems, respectively. This results in a set of synthetic sentence-pairs. Source sentences whose target-side is original are tagged with a special token ([Caswell et al., 2019](#)) (cf. Section 2.1). As in [Chinea-Ríos et al. \(2017\)](#), the original MT system is finally fine-tuned on these synthetic segment-pairs.

For mining monolingual sentences we create an efficient Trie structure given the large monolingual data. The idea is to store indices of the sentences (i.e. we restrict this number to 50) for each  $n$ -gram (upto trigram) of the corpus. Given the domain terms of the in-domain text, we can instantly retrieve the sentences from corpus.

## 2.4 Tuning Hyperparameters for Transformer

The NMT systems are Transformer models ([Vaswani et al., 2017](#)). To build our NMT systems, we used the MarianNMT ([Junczys-Dowmunt et al., 2018](#)) toolkit. The tokens of the training, evaluation and validation sets are segmented into sub-word units using Byte-Pair Encoding (BPE) ([Sennrich et al., 2016b](#)). Since English and Tamil are written in their own scripts and have no overlapping characters, BPE is applied individually on the source and target languages. Recently, [Sennrich and Zhang \(2019\)](#) demonstrated that commonly used hyperparameter configuration do not lead to the best results in low-resource settings. Accordingly, we carried out a series of experiments in order to find the best hyperparameter configuration for Trans-

former in our low-resource setting.<sup>4</sup> In particular, we played with some of the hyperparameters, and found that the following configuration lead to the best results in our low-resource translation settings: (i) the BPE vocabulary size: 8,000, (ii) the sizes of the encoder and decoder layers: 4 and 6, respectively, (iii) learning-rate: 0.0005, (iv) dropout ([Gal and Ghahramani, 2016](#)) between layers: 0.1. As for the remaining hyperparameters, we followed the recommended best setup from [Vaswani et al. \(2017\)](#). The models are trained with the Adam optimizer ([Kingma and Ba, 2014](#)), reshuffling the training corpora for each epoch. The early stopping criteria is based on cross-entropy; however, the final NMT system is selected as per the highest BLEU score on the validation set. The beam size for search is set to 12. We make our final NMT model with ensembles of 8 models that are sampled from the training run.

## 3 Experiments and Results

### 3.1 Data sets

This section presents the data sets that were used for system building. We used the monolingual and bilingual data provided by the WMT20 task organisers only. No external data has been used for the MT system building. Table 1 presents the corpus statistics. The parallel corpora released by

|                  | Parallel Data |            |            |
|------------------|---------------|------------|------------|
|                  | sentences     | words (EN) | words (TA) |
| train            | 350,142       | 6,489,872  | 5,763,047  |
| test             | 1,000         | 23,259     | 17,966     |
| dev.             | 989           | 23,415     | 17,901     |
| Monolingual Data |               |            |            |
| English          | 17M           | Tamil      | 31M        |

Table 1: The data statistics.

WMT20 for the English–Tamil task are from different sources (e.g. Tanzil v1<sup>5</sup> ([Tiedemann, 2012](#)), WikiMatrix<sup>6</sup> ([Schwenk et al., 2019](#)) and PMIndia<sup>7</sup> ([Haddow and Kirefu, 2020](#))). We merged segment-pairs of all data sources, and after applying standard cleaning scripts to the data we are left with

<sup>4</sup>This set of experiments were conducted on English-to-Tamil only and using the bitexts only. The best hyperparameter setup found in this task is used in the reverse translation task.

<sup>5</sup><http://opus.nlpl.eu/Tanzil-v1.php>

<sup>6</sup><https://ai.facebook.com/blog/wikimatrix/>

<sup>7</sup><http://data.statmt.org/pmindia>

<sup>3</sup><http://workshop.colips.org/news2016>

350K parallel segments (cf. row 3 of Table 1). As for the monolingual data, we used News-Crawl<sup>8</sup> and CommonCrawl<sup>9</sup> corpora (cf. last row of Table 1).

We observed that the corpora of one language (say, Tamil) contains sentences of other languages (e.g. English), so we use a language identifier<sup>10</sup> in order to remove such noise. In order to perform tokenisation for English and Tamil texts, we used the standard tool<sup>11</sup> in the Moses toolkit.

WMT20 released a development set of 1,989 sentences (*newsdev2020*) whose domain is naturally news. We used 1,000 sentences from *newsdev2020* as test set, and we call the test set *newstest1k*. The remaining sentences (989) are treated as the validation set.

### 3.2 The Baseline MT Systems

The BLEU scores of the NMT systems trained on the authentic parallel corpus (cf. Table 1) are reported in Table 2. These BLEU scores represent the MT systems that were trained following the best hyperparameter settings described in Section 2.4. Note that these MT systems serve our baselines. We refer the baseline MT system as Base. We see from Table 2 that the English-to-Tamil and

|                  | newstest1k |
|------------------|------------|
| English-to-Tamil | 5.81       |
| Tamil-to-English | 12.20      |

Table 2: The BLEU score of the baseline MT systems.

Tamil-to-English MT systems produce 5.81 and 12.20 BLEU scores, respectively, on the respective test sets. As expected, the translation quality from the morphologically-rich to morphologically-poor language improves.

### 3.3 The Improved MT Systems

We applied the pseudo in-domain sentence selection strategy described in Section 2.2 to the monolingual corpora (cf. Table 1), and considered the top-scored sentences for back-translation. Note that the in-domain language models for sentence

<sup>8</sup><http://data.statmt.org/news-crawl>

<sup>9</sup><http://data.statmt.org/news-discussions/>

<sup>10</sup>cld2: <https://github.com/CLD2Owners/cld2>

<sup>11</sup><https://github.com/moses-smt/mosesdecoder/blob/master/scripts/tokenizer/tokenizer.perl>

selection were built on the PMI data whose domain is news in nature. The development set sentences (998 sentences; cf. Table 2) were also appended to the PMI data. The BLEU scores on the MT systems trained on the augmented training data are presented in Table 3. We stopped the iterative train-

|                  | newstest1k<br>BLEU | newstest2020<br>SacreBLEU |
|------------------|--------------------|---------------------------|
| English-to-Tamil |                    |                           |
| Base+ 800K       | 8.12               |                           |
| Base+ 1.5M       | 9.35               |                           |
| Base+ 2.7M       | 9.45               | 5.4                       |
| Tamil-to-English |                    |                           |
| Base + 800K      | 18.33              |                           |
| Base + 1.5M      | 18.52              |                           |
| Base + 2.7M      | 19.41              |                           |
| Base + 3.3M      | 19.91              | 14.7                      |

Table 3: The BLEU scores of the MT systems trained on the interactively augmented training data.

ing process (cf. Section 2.1) when there were no significant improvements in terms of the test set BLEU scores. This training process provides us with the improved MT systems. As can be seen from Table 3, the final MT systems surpass the respective baseline MT systems with large margins.

We translate the blind test sets (*newstest2020*) for the English-to-Tamil and Tamil-to-English translation tasks released by WMT20 by the best MT systems (cf. Table 3). The blind test sets for the English-to-Tamil and Tamil-to-English tasks contain 6,988 and 997 segments, respectively. The sacreBLEU (Post, 2018) scores of the best NMT systems on *newstest2020* are shown in the last column of Table 3.<sup>12</sup>

### 3.4 Fine-tuning the best NMT systems

This section presents the MT systems that were prepared by the adaptation technique described in Section 2.3. We mine the source and target monolingual sentences from the large monolingual corpora given the terms and NEs (and transliterated NEs) extracted from *newstest1k*.<sup>13</sup> As described in Section 2.3, synthetic data is created by translating the source and target sentences by the target-to-source and source-to-target MT systems (cf. Table 3; the best MT systems), respectively. Finally, the

<sup>12</sup>The SacreBLEU scores were taken from OCELOT <https://ocelot.mteval.org>

<sup>13</sup>Note that NEs were extracted from the English text only.

best MT system is fine-tuned on the synthetic data. The BLEU scores of the adapted MT systems on newstest1k are reported in Section 4. When we compare the original MT systems reported in Table 3 with the adapted MT systems, we see that (i) the English-to-Tamil adapted MT system produces a 1.55 BLEU points (corresponding to 16.4% relative) improvement over the the original English-to-Tamil MT system, and (ii) the Tamil-to-English adapted MT system produces a 1.41 BLEU points (corresponding to 7.08% relative) improvement over the the original Tamil-to-English MT system. The improvements are statistically significant.

|                  | newstest1k<br>BLEU | newstest2020<br>SacreBLEU |
|------------------|--------------------|---------------------------|
| English-to-Tamil | 10.80              | 6.1                       |
| Tamil-to-English | 21.32              | 15.8                      |

Table 4: The BLEU and SacreBLEU scores of the adapted MT systems on newstest1k and newstest2020, respectively.

As above, we create the English-to-Tamil and Tamil-to-English adapted MT systems for the blind test sets. Then, we translate the blind test sets with the adapted MT systems. The sacreBLEU (Post, 2018) scores of the adapted MT systems on newstest2020 are shown in the last column of Table 4. Again, the adaption strategy brings about moderate improvements over the original MT systems, i.e. a 0.7 SacreBLEU points (corresponding to 13% relative) improvement for the English-to-Tamil translation and a 1.1 SacreBLEU points (corresponding to 7.5% relative) improvement for the Tamil-to-English translation.

## 4 Conclusion

This paper presents the ADAPT system description for the WMT20 News Translation Shared Task. We participated in the English-to-Tamil and Tamil-to-English tasks. English–Tamil is a low-resource language-pair and we used the data provided by the WMT20 organisers only. Given the limited resources provided for the tasks, we aimed to build the competitive translation systems for the competition. For this, we applied a variety of strategies, e.g. iterative data augmentation, selection of pseudo in-domain sentences, and a novel strategy for the adaptation of the NMT models to the task. We found that the systematic addition of these techniques to baseline yields excellent improvements

over the baseline.

This paper presented an effective adaptation method for the NMT systems. This method is found to be effective as far as the translation task we participated in is concerned. In the future, we aim to test on-the-fly adaptation method (Farajian et al., 2017, 2018) to translate domain texts.

## Acknowledgments

The ADAPT Centre for Digital Content Technology is funded under the Science Foundation Ireland (SFI) Research Centres Programme (Grant No. 13/RC/2106) and is co-funded under the European Regional Development Fund. This project has partially received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 713567, and the publication has emanated from research supported in part by a research grant from SFI under Grant Number 13/RC/2077.

## References

- Amittai Axelrod, Xiaodong He, and Jianfeng Gao. 2011. [Domain adaptation via pseudo in-domain data selection](#). In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 355–362, Edinburgh, Scotland, UK. Association for Computational Linguistics.
- Loïc Barrault, Ondřej Bojar, Marta R. Costa-jussà, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Philipp Koehn, Shervin Malmasi, Christof Monz, Mathias Müller, Santanu Pal, Matt Post, and Marcos Zampieri. 2019. [Findings of the 2019 conference on machine translation \(wmt19\)](#). In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 1–61, Florence, Italy. Association for Computational Linguistics.
- Nikolay Bogoychev and Rico Sennrich. 2019. Domain, translationese and noise in synthetic data for neural machine translation. *arXiv preprint arXiv:1911.03362*.
- Franck Burlot and François Yvon. 2018. [Using monolingual data in neural machine translation: a systematic study](#). In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 144–155, Belgium, Brussels. Association for Computational Linguistics.
- Isaac Caswell, Ciprian Chelba, and David Grangier. 2019. [Tagged back-translation](#). In *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, pages 53–63, Florence, Italy. Association for Computational Linguistics.

- Peng-Jen Chen, Jiajun Shen, Matthew Le, Vishrav Chaudhary, Ahmed El-Kishky, Guillaume Wenzek, Myle Ott, and Marc’Aurelio Ranzato. 2019. Facebook AI’s WAT19 Myanmar-English translation task submission. In *Proceedings of the 6th Workshop on Asian Translation*, pages 112–122, Hong Kong, China.
- Mara China-Ríos, Álvaro Peris, and Francisco Casacuberta. 2017. [Adapting neural machine translation with parallel synthetic data](#). In *Proceedings of the Second Conference on Machine Translation*, pages 138–147, Copenhagen, Denmark. Association for Computational Linguistics.
- Xiangyu Duan, Rafael Banchs, Min Zhang, Haizhou Li, and A. Kumaran. 2016. [Report of NEWS 2016 machine transliteration shared task](#). In *Proceedings of the Sixth Named Entity Workshop*, pages 58–72, Berlin, Germany. Association for Computational Linguistics.
- M Amin Farajian, Nicola Bertoldi, Matteo Negri, Marco Turchi, and Marcello Federico. 2018. Evaluation of terminology translation in instance-based neural mt adaptation. In *Proceedings of the European Association for Machine Translation*.
- M Amin Farajian, Marco Turchi, Matteo Negri, and Marcello Federico. 2017. Multi-domain neural machine translation through unsupervised adaptation. In *Proceedings of the Second Conference on Machine Translation*, pages 127–137.
- Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. [Incorporating non-local information into information extraction systems by Gibbs sampling](#). In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL’05)*, pages 363–370, Ann Arbor, Michigan. Association for Computational Linguistics.
- George Foster, Cyril Goutte, and Roland Kuhn. 2010. Discriminative instance weighting for domain adaptation in statistical machine translation. In *Proceedings of the 2010 conference on empirical methods in natural language processing*, pages 451–459.
- Yarin Gal and Zoubin Ghahramani. 2016. [A theoretically grounded application of dropout in recurrent neural networks](#). *CoRR*, abs/1512.05287.
- Jianfeng Gao, Joshua Goodman, Mingjing Li, and Kai-Fu Lee. 2002. Toward a unified approach to statistical language modeling for chinese. *ACM Transactions on Asian Language Information Processing (TALIP)*, 1(1):3–33.
- Barry Haddow and Faheem Kirefu. 2020. Pmindia—a collection of parallel corpora of languages of india. *arXiv preprint arXiv:2001.09907*.
- Rejwanul Haque, Sandipan Dandapat, Ankit Kumar Srivastava, Sudip Kumar Naskar, and Andy Way. 2009. [English-Hindi transliteration using context-informed PB-SMT: the DCU system for NEWS 2009](#). In *Proceedings of the 2009 Named Entities Workshop: Shared Task on Transliteration (NEWS 2009)*, pages 104–107, Suntec, Singapore. Association for Computational Linguistics.
- Rejwanul Haque, Mohammed Hasanuzzaman, and Andy Way. 2020. Analysing terminology translation errors in statistical and neural machine translation. *Machine Translation (in press)*, 34.
- Rejwanul Haque, Sergio Penkale, and Andy Way. 2014. Bilingual termbank creation via log-likelihood comparison and phrase-based statistical machine translation. In *Proceedings of the 4th International Workshop on Computational Terminology (Computerm)*, pages 42–51, Dublin, Ireland.
- Rejwanul Haque, Sergio Penkale, and Andy Way. 2018. [TermFinder: log-likelihood comparison and phrase-based statistical machine translation models for bilingual terminology extraction](#). *Language Resources and Evaluation*, 52(2):365–400.
- Vu Cong Duy Hoang, Philipp Koehn, Gholamreza Haffari, and Trevor Cohn. 2018. [Iterative back-translation for neural machine translation](#). In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, pages 18–24, Melbourne, Australia. Association for Computational Linguistics.
- Matthias Huck, Viktor Hangya, and Alexander Fraser. 2019. Better oov translation with bilingual terminology mining. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5809–5815.
- Marcin Junczys-Dowmunt, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann, Alham Fikri Aji, Nikolay Bogoychev, André F. T. Martins, and Alexandra Birch. 2018. [Marian: Fast neural machine translation in C++](#). In *Proceedings of ACL 2018, System Demonstrations*, pages 116–121, Melbourne, Australia. Association for Computational Linguistics.
- Diederik P. Kingma and Jimmy Ba. 2014. [Adam: A method for stochastic optimization](#). *CoRR*, abs/1412.6980.
- Alberto Poncelas, Dimitar Shterionov, Andy Way, Gideon Maillette de Buy Wenniger, and Peyman Passban. 2018. Investigating backtranslation in neural machine translation. In *Proceedings of The 21st Annual Conference of the European Association for Machine Translation (EAMT 2018)*, pages 249–258, Alicante, Spain.
- Matt Post. 2018. [A call for clarity in reporting BLEU scores](#). In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Belgium, Brussels. Association for Computational Linguistics.

- Paul Rayson and Roger Garside. 2000. Comparing corpora using frequency profiling. In *The workshop on comparing corpora*, pages 1–6.
- Holger Schwenk, Vishrav Chaudhary, Shuo Sun, Hongyu Gong, and Francisco Guzmán. 2019. Wikimatrix: Mining 135m parallel sentences in 1620 language pairs from wikipedia. *arXiv preprint arXiv:1907.05791*.
- 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 (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany. Association for Computational 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 (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany.
- Rico Sennrich and Biao Zhang. 2019. [Revisiting low-resource neural machine translation: A case study](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 211–221, Florence, Italy. Association for Computational Linguistics.
- Jörg Tiedemann. 2012. Parallel data, tools and interfaces in OPUS. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'2012)*, pages 2214–2218, Istanbul, Turkey.
- Antonio Toral. 2013. Hybrid selection of language model training data using linguistic information and perplexity. In *Proceedings of the second workshop on hybrid approaches to translation*, pages 8–12.
- Nicola Ueffing, Gholamreza Haffari, and Anoop Sarkar. 2007. [Transductive learning for statistical machine translation](#). In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 25–32, Prague, Czech Republic. Association for Computational Linguistics.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 6000–6010.
- Joern Wuebker, Patrick Simianer, and John DeNero. 2018. Compact personalized models for neural machine translation. *arXiv preprint arXiv:1811.01990*.
- Keiji Yasuda, Ruiqiang Zhang, Hirofumi Yamamoto, and Eiichiro Sumita. 2008. Method of selecting training data to build a compact and efficient translation model. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-II*.
- Jiajun Zhang and Chengqing Zong. 2016. [Exploiting source-side monolingual data in neural machine translation](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545, Austin, Texas. Association for Computational Linguistics.