HW-TSC's Participation in the WMT 2021 News Translation Shared Task

Daimeng Wei, Zongyao Li, Zhanglin Wu, Zhengzhe Yu,
Xiaoyu Chen, Hengchao Shang, Jiaxin Guo, Minghan Wang,
Lizhi Lei, Min Zhang, Hao Yang, Ying Qin,
Huawei Translation Service Center, Beijing, China
{weidaimeng,lizongyao,wuzhanglin2,yuzhengzhe,
chenxiaoyu35,shanghengchao,guojiaxin1,wangminghan,
leilizhi,zhangmin186,yanghao30,qinying}@huawei.com

Abstract

This paper presents the submission of Huawei Translate Services Center (HW-TSC) to the WMT 2021 News Translation Shared Task. We participate in 7 language pairs, including Zh/En, De/En, Ja/En, Ha/En, Is/En, Hi/Bn, and Xh/Zu in both directions under the constrained condition. We use Transformer architecture and obtain the best performance via multiple variants with larger parameter sizes. We perform detailed pre-processing and filtering on the provided large-scale bilingual and monolingual datasets. Several commonly used strategies are used to train our models, such as Back Translation, Forward Translation, Multilingual Translation, Ensemble Knowledge Distillation, etc. Our submission obtains competitive results in the final evaluation.

1 Introduction

This paper introduces our submission to the WMT 2021 News Translation Shared Task. We participate in seven language pairs including Chinese/English (Zh/En), German/English (De/En), Japanese/English (Ja/En), Hausa/English (Ha/En), Icelandic/English (Is/En), Hindi/Bengali (Hi/Bn), and Xhosa/Zulu (Xh/Zu) in both directions. We consider that the officially provided dataset has the acceptable size and quality and therefore only participate in the constrained evaluation. Our method is mainly based on previous works but with fine-grained data cleansing techniques and language-specific optimizations.

For each language pair, we perform multi-step data cleansing on the provided dataset and only keep a high-quality subset for training. At the same time, several strategies are tested in a pipeline, including Backward (Edunov et al., 2018) and Forward(Wu et al., 2019a) Translation, Multilingual Translation (Johnson et al., 2017), Right-to-Left Models (Liu et al., 2016), Iterative Joint Training (Zhang et al., 2018), Ensemble Knowledge Distillation (Freitag et al., 2017; Li et al., 2019) , Fine-Tuning (Sun et al., 2019), Ensemble (Garmash and Monz, 2016), and PostProcess.

We combined all the techniques mentioned above and the overall training process is shown in Figure 1. Section 2 focuses on our data processing strategies while section 3 describes our training techniques, including model architecture and iterative training, etc. Section 4 explains our experiment settings and training processes and section 5 presents our experiment results.

2 Data

2.1 Data Source

For all language pairs, we follow the constrained data requirements and take full advantages of the bilingual and monolingual training data available. Table 1 lists the data sizes of each language pair before and after filtering.

2.2 Data Pre-processing

We use following operations to pre-process the data:

- Filter out repeated sentences (Khayrallah and Koehn, 2018; Ott et al., 2018).
- Convert XML escape characters.
- Normalize punctuations using Moses (Koehn et al., 2007).
- Delete html tags, non-UTF-8 characters, unicode characters and invisible characters.
- Filter out sentences with mismatched parentheses and quotation marks; sentences of which punctuation percentage exceeds 0.3; sentences with the character-to-word ratio greater than 12 or less than 1.5; sentences of which the source-to-target token ratio higher

Proceedings of the Sixth Conference on Machine Translation (WMT), pages 225–231
November 10–11, 2021. ©2021 Association for Computational Linguistics

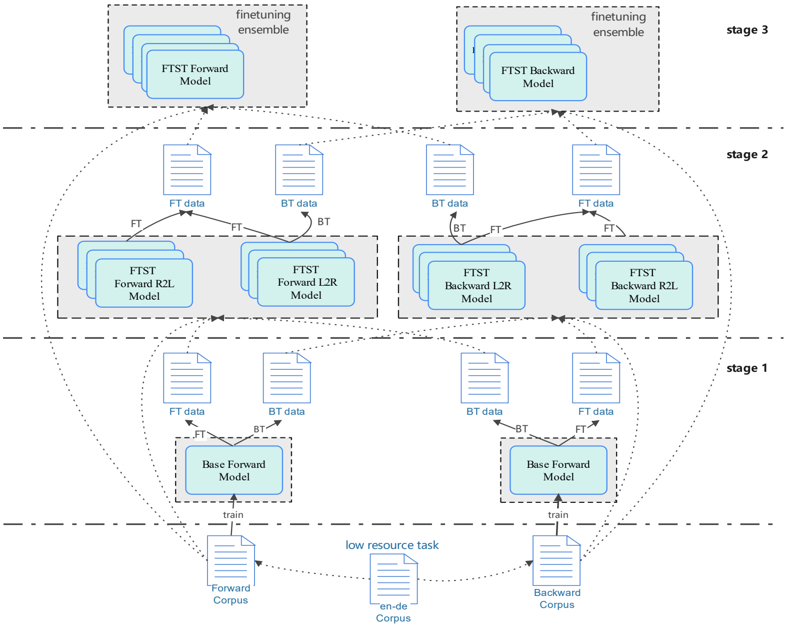

Figure 1: This figure shows the training process for the WMT 2021 News Translation Shared Task, which consists of three stages. In stage 1, one forward model and one backward model are trained. In stage 2, the synthetic data by FTST is used to train L2R and R2L models. In stage 3, the synthetic data by enhanced models are used to train models. Finally, model ensemble is used to boost the performance.

than 3 or lowers than 0.3; sentences with more than 120 tokens.

- Apply langid (Joulin et al., 2016b,a) to filter sentences in other languages.
- Use fast-align (Dyer et al., 2013) to filter sentence pairs with poor alignment.

We perform the additional steps to process Chinese data:

- Convert traditional Chinese characters to simplified ones.
- Convert fullwidth forms to halfwidth forms.

Data sizes before and after cleansing are listed in Table 1.

2.3 Data Selection

Since the news (in-domain) monolingual data in some tasks is not sufficient, it is necessary to obtain data from Common Crawl. We use Fasttext (Joulin et al., 2016a) to train a binary classification model to distinguish between in-domain and out-domain data.

3 System Overview

3.1 Model

Transformer (Vaswani et al., 2017) has been widely used for machine translation in recent years, which has achieved good performance even with the most primitive architecture without much modifications. Therefore, we choose to start from Transformer-Big and consider it as a baseline. Four variants of Transformer are also evaluated during the experiments, which are the model with wider FFN layers proposed in (Ng et al., 2019), and the deeper encoder version proposed in (Sun et al., 2019). Here, we use the following four variants:

- Deep 25-6 model: The number of the encoder layers is adjusted to 25 based on the transformer base model architecture and layer normalization is added. The other settings remain the same as the base model.
- Deep 35-6 model: The number of the encoder layers is adjusted to 36 based on the transformer base model architecture and layer normalization is added. The other settings remain the same as the base model.

language pairs	Raw bi data	Filter bi data	Used mono data
Zh/En	37.8M	16.5M	En: 150M, Zh:150M
De/En	95M	79M	En: 230M, De: 317M
Ja/En	18M	13.5M	En: 300M, Ja: 300M
Ha/En	0.73M	0.59M	En: 8M, Ha:8.65M
Is/En	5.69M	4.04M	En: 20M, Is: 18M
Hi/Bn	3.53M	3.4M	Bn: 59.3M, Hi: 45.8M

Table 1: Bilingual data sizes before and after filtering, and monolingual data used in tasks.

- Deep 35-6 big model: This model features 35-layer encoder, 6-layer decoder, 768 dimensions of word vector, 3076 dimensions of FFN, 16-head self-attention, and pre-norm.

- Deep 25-6 large Model: This model features 25-layer encoder, 6-layer decoder, 1024 dimensions of word vector, 4096 dimensions of FFN, 16-head self-attention, and pre-norm.

3.2 Data Augmentation

Back-translation (Edunov et al., 2018) is an effective way to boost translation quality by using monolingual sentences to generate synthetic training parallel data. As described in (Wu et al., 2019b), similar to back translation, the monolingual corpus in source language can also be used to generate forward translation text with a trained MT model, and the generated forward and backward translation data can both be merged with the authentic bilingual data. This strategy can increase the data size to a large extent.

We take full advantages of the officially provided monolingual data for data augmentation. In terms of back translation, we adopt top-k sampling for high-resource languages, and adopt beam search for low-resource languages. With regard to forward translation, we translate monolingual data using beam search. Through sampling, we ensure that the sizes of data generated by forward and back translation are relatively equal. In this paper, we refer to the combination of forward and sampling back translation as FTST.

3.3 Iterative Joint Training

Zhang et al. (2018) propose a new iterative joint training method, that is, using monolingual data from both source and target sides to train a source-to-target (forward) model and a target-to-source (backward) model at the same time. The two models generate synthetic data for each other. The advantage of such method is that both of the two models gain improvement after each iteration with the synthetic data provided by the other, and then can generate synthetic data with higher quality. Such training procedure is repeated after the two models converge.

3.4 Multilingual Translation

Johnson et al. (2017) propose a simple solution to use a single neural machine translation model to translate among multiple languages, and the model requires no change to the model architecture. Instead, the model introduces an artificial token at the beginning of the input sentence to specify the required target language. All languages use a shared vocabulary. There is no need to add more parameters. In low-resource tasks, we select a portion of the En-De bilingual data and conduct a joint training. The experiment shows that a multilingual model can improve the translation quality of low-resource languages to a large extent.

3.5 Right-to-Left Models

The approach of Right-to-Left is proposed by (Liu et al., 2016). The main idea is to integrate information of Right-to-Left (R2L) models to Left-to-Right (L2R) ones. Following this strategy, we translate the source sentences of the monolingual data with both R2L models and L2R models. In the Zh/En and De/En tasks, we use the R2L model to synthesize forward translation data using beam search and mix the synthetic data with the L2R synthetic data for iterative joint training.

3.6 Ensemble Knowledge Distillation

Ensemble Knowledge Distillation (Freitag et al., 2017; Li et al., 2019) improves the performance of a student model by distilling knowledge from a group of trained teacher models. Comparing with some soft label distillation methods, the EKD for NMT is relatively straightforward, which can be implemented by training the student models on the combination of the original training set and the translation from the ensembled teacher model on the training set. In our experiments, we ensemble models as the teacher model to translate the wmt21 test set, and use the translate results to further fine-tune models.

3.7 Fine-tuning

Previous works have demonstrated that fine-tuning a model with in-domain data, such as last year's test set, could effectively improve the performance of this year (Sun et al., 2019). We use the dev and test sets from previous years, coupled with data generated by ensemble knowledge distillation and noises added to the target side, to fine-tune models and achieve further improvements.

3.8 Ensemble

Model ensemble is a widely used technique in previous WMT workshops (Garmash and Monz, 2016), which can improve the performance by combining the predictions of several models at each decoding step. In our work, we ensemble models with different architectures to further improve system performances. For Zh/En and De/En, we experimented with a combination of the Deep 35-6 big model and the Deep 25-6 large model to ensemble. For all language pairs, we train multiple models to ensemble by shuffle the data.

4 Experiment Settings

4.1 Settings

We use the open-source fairseq (Ott et al., 2019) for training and sacreBLEU (Post, 2018) to measure system performances. The main parameters are as follows: Each model is trained using 8 GPUs. The size of each batch is set as 2048, parameter update frequency as 32, and learning rate as 5e-4 (Vaswani et al., 2017). The number of warmup steps is 4000, and model is saved every 1000 steps. The architectures we used are described in section 3.1. We adopt dropout, and the rate varies across different language pairs. Marian (Junczys-Dowmunt et al., 2018) is used for decoding during inference.

4.2 Training Process

We employ iterative training and phase-based data augmentation. Figure 1 shows our training process in details. The specific steps are as follows:

1) Process data using methods described in section 2.2. Train one forward model and one backward model.

System	en2zh	zh2en
baseline	39.1	26.5
FTST	45.1 (+6.0)	32.4 (+5.9)
in-domain FTST + R2L	46.2 (+1.1)	34.4 (+2.0)
finetuning	46.5 (+0.3)	34.8 (+0.4)
ensemble	46.7 (+0.2)	34.9 (+0.1)
wmt21 final submit	35.1	28.9

Table 2: The experimental result of Zh/En tasks

2) Generate back translation and forward translation data. Mix the data with parallel training data and train three forward L2R models and three backward models. At the same time, train three R2L models for generating R2L forward translation data, in order to improve the diversity of synthetic data.

3) Split monolingual data into several sets. Generate back translation and forward translation data using models trained in step 2. Mix sampled synthetic data with bilingual training data and train four forward models and four backward models.

4) Average the last five checkpoints of each model and fine-tune it. Ensemble models to produce the final system.

5 Results and analysis

5.1 Zh/En

We use methods described in Section 2.2 for data processing. Four model architectures mentioned in Section 3.1 are employed to increase system diversity. On the basis of bilingual baselines model, we use FTST data augmentation to further enhance model performance.

Table 2 lists the results of our submission on WMT 2020 News Task test set. Comparing with the baseline model, FTST leads to 6.0 BLEU increase on en2zh direction and 5.9 BLEU increase on the opposite direction. We conduct data distillation on source sentences from WMT 2017 and 2018 news test sets, mix the generated data with the original data, and add noises to the target side. We fine-tune the model using the mixed data and achieve 1.1 BLEU and 2.0 BLEU increases on en2zh and zh2en directions, respectively. We then conduct a second-round FTST data augmentation on the fine-tuned model. In this round, we adopt the R2L model. We conduct data distillation on source sentences from WMT 2017-2018 news test sets, mix

System	en2de	de2en
baseline	33.1	39.7
FTST	34.2 (+1.1)	40.8 (+1.1)
FTST + R2L	34.5 (+0.3)	41.1 (+0.3)
finetuning	38.2 (+3.7)	43.1 (+2.0)
ensemble	38.3 (+0.1)	43.4 (+0.3)
postprocess	39.7 (+1.4)	-
wmt21 final submit	29.8	34.7

Table 3: The experimental result of De/En tasks

System	en2ja	ja2en
baseline	36.4	21.4
iterative FTST	39.2 (+2.8)	23.1 (+2.7)
finetuning	42.9 (+3.7)	25.3 (+2.2)
ensemble	43.6 (+0.7)	26.0 (+0.7)
wmt21 final submit	45.4	26.5

Table 4: The experimental result of Ja/En tasks

the generated data with the WMT 2017-2018 test sets, and add noises to the target side. We fine-tune the model using the mixed data and achieve 0.3 BLEU and 0.4 BLEU increases on en2zh and zh2en directions, respectively. Finally, ensemble further leads to 0.2 BLEU increase on the en2zh direction and 0.1 BLEU increase on the opposite direction. When submitting the final results, we further fine-tune the model with WMT 2019 and 2020 test sets. Our models achieve 35.1 BLEU on the en2zh direction and 28.9 BLEU on the zh2en direction when measuring with the WMT 2021 News Task test set.

5.2 De/En

For the En-De task, we adopt the Deep 36-5 big model and Deep 25-6 large model, as described in section 3.1. We use Moses for English and German word segmentation. The training data are segmented by a shared SentencePiece model. The source and target side each has a vocabulary with 32K words. We process all data using filter methods described in section 2.2.

Table 3 lists the results of our submission on WMT 2020 News Task test set. Comparing with the baseline model, two rounds of FTST data augmentation contribute to 1.4 BLEU increase on each directions. We conduct data distillation on source sentences from WMT 2020 news test sets, mix the generated data with the WMT 2018 and WMT 2019 test sets after adding noises to the target side. We fine-tune the model using the mixed data and achieve 3.7 BLEU and 2.0 BLEU increases on en2de and de2en directions, respectively. Ensemble further leads to 0.1 BLEU increase on the en2de direction and 0.3 BLEU increase on the opposite direction. Ensemble does not have significant impact on this task. It should be noted that we find that the quotation marks generated by the en2de model does not comply with the German standard, so we add a correction script to the post-processing(just convert English quotation marks to German quotation marks), which surprisingly leads to 1.4 BLEU increase. When submitting the final results, we further fine-tune the model with WMT 2020 test set. Our submitted models achieve 29.8 BLEU on the en2de direction and 34.7 BLEU on the de2en direction when measuring with the WMT 2021 News Task test set.

5.3 Ja/En

For Ja/En task, we adopt the same settings as that for the Zh-En task. The dropout rate is set to 0.1. The training data are segmented by a shared SentencePiece model. The source and target side each has a vocabulary with 32K words. The size of parallel data after cleansing is 13.5M. We sampled 150M English monolingual data from News Crawl and 300M Japanese monolingual data from News Crawl and Common Crawl (150M from each source).

Table 4 lists the results of our submission on WMT 2020 News Task test set. Comparing with the baseline model, iterative FTST data augmentation contribute to 2.8 BLEU and 1.7 BLEU increases on the en2ja and ja2en directions respectively. We conduct data distillation on source sentences from WMT 2020 news test sets, mix the generated data with the WMT 2020 dev set after adding noises to the target side. We fine-tune the model using the mixed data and achieve 3.7 BLEU and 2.2 BLEU increases on en2ja and ja2en directions, respectively. We train four models on each direction and ensemble further leads to 0.9 BLEU increase on the en2ja direction and 1.0 BLEU increase on the opposite direction. When submitting the final results, we further fine-tune the model with WMT 2021 dev set. Our submitted models achieve 45.4 BLEU on the en2ja direction and 26.5 BLEU on the j2en direction when measuring with the WMT 2021 News Task test set.

System	en2ha	ha2en	en2is	is2en	hi2bn	bn2hi	xh2zu	zu2xh
baseline	2.8	7.7	18.3	25.1	7.4	18.0	2.1	6.2
multilingual (add en2de data)	14.9	18.9	20.2	28.0	9.2	18.3	7.3	8.1
iFTBT	19.7	23.2	23.5	32.4	10.4	19.4	9.3	9.2
wmt21 final submit	20.3	17.5	27.5	38.4	13.0	21.9	11.8	9.9

Table 5: The experimental result of low resource tasks. iBTFT indicates that multiple rounds of BTFT are used for data enhancement.

5.4 Low resource tasks

We use the same strategy to deal with low resource tasks (En-Ha, En-Is, Bn-Hi and Xh-Zu). We train a bilingual baseline model and a monolingual baseline model for each direction. Every multilingual model is trained with 10x bilingual data sampled from the training corpora and 50M En-De parallel data. For en2ha, en2is, hi2bn and xh2zu, we use en2de data for training. For other language directions, we use de2en data for training.

Table 5 lists the results of our submission on dev set. On the eight language directions, all multilingual models gain huge improvements when comparing with the bilingual baseline model. Particularly, En-Ha achieves the greatest improvements: 12.1 BLEU on en2ha direction and 11.20 on ha2en direction. Bn-Hi achieves the slightest improvements: 0.4 BLEU on bn2hi direction and 1.78 on hi2bn direction. The results demonstrate that the fewer the bilingual data, the greater impact a multilingual model has. In other extremely low-resource scenarios, the improvement gained by a multilingual model for En-Ha is greater than that for the Xh-Zu task. We think the reason lies in the differences of language similarities. On the basis of multilingual models, we conduct data augmentation as described in section 3.2. We adjust sampling ratios according to the monolingual data size of each languages. Our data augmentation strategy achieves improvements on all eight language directions, from 1.1 BLEU to 4.4 BLEU increase. When conduct the second-round FTST data augmentation, we only get a slight increase on the En-Ha task: 0.2 BLEU on en2ha direction and 0.9 on ha2en direction. We also leverage fine-tuning and ensemble techniques to further improve our model performances. Finally, we get the highest BLEU score on the xh2en direction and the second highest BLEU score on the en2ha direction.

6 Conclusion

This paper presents the submissions of HW-TSC to the WMT 2021 News Translation Task. For each direction in all pairs, we perform experiments with a series of pre-processing and training strategies. The effectiveness of each strategy is demonstrated. Our experiments show that in low-resource scenarios, multilingual model that utilizing data from other languages can improve system performance to a large extent. Data augmentation strategy is still effective for multilingual models. Our submissions finally achieves competitive results in the evaluation.

References

Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. A simple, fast, and effective reparameterization of IBM model 2. In *Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, June 9-14, 2013, Westin Peachtree Plaza Hotel, Atlanta, Georgia, USA*, pages 644–648.

Sergey Edunov, Myle Ott, Michael Auli, and David Grangier. 2018. Understanding back-translation at scale. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 489–500.

Markus Freitag, Yaser Al-Onaizan, and Baskaran Sankaran. 2017. Ensemble distillation for neural machine translation. *CoRR*, abs/1702.01802.

Ekaterina Garmash and Christof Monz. 2016. Ensemble learning for multi-source neural machine translation. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1409–1418.

Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hérve Jégou, and Tomas Mikolov. 2016a. Fasttext.zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651*.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016b. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.

Marcin Junczys-Dowmunt, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann, Alham Fikri Aji, Nikolay Bogoychev, et al. 2018. Marian: Fast neural machine translation in c++. *arXiv preprint arXiv:1804.00344*.

Huda Khayrallah and Philipp Koehn. 2018. On the impact of various types of noise on neural machine translation. *arXiv preprint arXiv:1805.12282*.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic. Association for Computational Linguistics.

Bei Li, Yinqiao Li, Chen Xu, Ye Lin, Jiqiang Liu, Hui Liu, Ziyang Wang, Yuhao Zhang, Nuo Xu, Zeyang Wang, Kai Feng, Hexuan Chen, Tengbo Liu, Yanyang Li, Qiang Wang, Tong Xiao, and Jingbo Zhu. 2019. The niutrans machine translation systems for WMT19. In *Proceedings of the Fourth Conference on Machine Translation, WMT 2019, Florence, Italy, August 1-2, 2019 - Volume 2: Shared Task Papers, Day 1*, pages 257–266.

Lemao Liu, Masao Utiyama, Andrew Finch, and Eiichiro Sumita. 2016. Agreement on target-bidirectional neural machine translation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 411–416.

Nathan Ng, Kyra Yee, Alexei Baevski, Myle Ott, Michael Auli, and Sergey Edunov. 2019. Facebook FAIR's WMT19 News Translation Task Submission. *arXiv preprint arXiv:1907.06616*.

Myle Ott, Michael Auli, David Grangier, and Marc'Aurelio Ranzato. 2018. Analyzing uncertainty in neural machine translation. In *International Conference on Machine Learning*, pages 3956–3965. PMLR.

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. *arXiv preprint arXiv:1904.01038*.

Matt Post. 2018. A call for clarity in reporting bleu scores. *arXiv preprint arXiv:1804.08771*.

Meng Sun, Bojian Jiang, Hao Xiong, Zhongjun He, Hua Wu, and Haifeng Wang. 2019. Baidu neural machine translation systems for WMT19. In *Proceedings of the Fourth Conference on Machine Translation, WMT 2019, Florence, Italy, August 1-2, 2019 - Volume 2: Shared Task Papers, Day 1*, pages 374–381.

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 5998–6008.

Lijun Wu, Yiren Wang, Yingce Xia, Tao Qin, Jianhuang Lai, and Tie-Yan Liu. 2019a. Exploiting monolingual data at scale for neural machine translation. 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 4205–4215.

Lijun Wu, Yiren Wang, Yingce Xia, Tao Qin, Jianhuang Lai, and Tie-Yan Liu. 2019b. Exploiting monolingual data at scale for neural machine translation. 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)*, pages 4207–4216.

Zhirui Zhang, Shujie Liu, Mu Li, Ming Zhou, and Enhong Chen. 2018. Joint training for neural machine translation models with monolingual data. In *Thirty-Second AAAI Conference on Artificial Intelligence*.