Improving Chinese-centric Low-resource Translation Using English-centric Pivoted Parallel Data

Xiaoying Wang
Zhengzhou University
Henan, China
0009-0002-8870-3095

Lingling Mu
Zhengzhou University
Henan, China
iellmu@zzu.edu.cn

Hongfei Xu
Zhengzhou University
Henan, China
0000-0001-8397-1459

***Abstract*—The good performance of Neural Machine Translation (NMT) normally relies on a large amount of parallel data, while the bilingual data between languages are usually insufficient. mBART improves the performance of low-resource translation by pre-training on multilingual monolingual data and then fine-tuning on bilingual data, but does not leverage parallel data which contains crucial alignment information between languages. In this paper, we propose to use English-centric parallel data in a Multilingual NMT (MNMT) manner with English as the pivot, to provide translation and alignment information for the translation between Chinese and other languages. We conduct experiments on the CCMT 2023 low-resource machine translation task between Chinese and the languages among "the Belt and Road". Our method improves the zh→vi, vi→zh, zh→mn, mn→zh, zh→cs and cs→zh tasks by $+1.65$, $+0.24$, $+0.91$, $+3.47$, $+2.88$, $+6.35$ BLEU respectively over the strong mBART baseline, showing the effectiveness of our approach and the importance of English-centric parallel data.**

***Index Terms*—MNMT, Low-resource Translation, Pre-trained Models**

I. Introduction

Neural Machine Translation (NMT) [6], [49], [53] can deliver high quality translations with the support of large amounts of high quality parallel data, but it normally performs poorly on low-resource tasks. However, the amount of bilingual data between non-English languages is normally insufficient.

Pre-training [12], [29] and then fine-tuning the pre-trained models has been proven effective in improving the performance in low-resource scenarios. In terms of low-resource translation, current methods [33], [42] first pre-train on large-scale monolingual data of multiple languages in a de-noising manner, and then fine-tuning on the parallel data of the target task. This usually can lead to significant improvements by leveraging the language understanding/generation knowledge gained during large scale pre-training.

However, the pre-training stage lacks alignment information between languages which is crucial for translation. While the parallel data between Chinese and other languages is not sufficient, there might be large scale datasets between these languages and English, and leveraging such data may bring in crucial knowledge for alignment and translation using English as the pivot. In this paper, we propose to leverage English-centric parallel data for the low-resource translation between Chinese and other languages in a Multilingual NMT (MNMT) manner. We test our approach on the CCMT 2023 low-resource machine translation task between Chinese and the languages among "the Belt and Road", and obtain significant improvements over the strong mBART baseline, showing the effectiveness of our approach and the importance of using English-centric parallel data.

Our main contributions are as follows:

- We propose to leverage English-centric parallel data to improve the performance of Chinese-centric Low-resource translation in an MNMT manner.
- Our experiments on the CCMT 2023 low-resource translation tasks between Chinese and the languages among "the Belt and Road" leads to $+1.65$, $+0.24$, $+0.91$, $+3.47$, $+2.88$, $+6.35$ BLEU improvements for zh→vi, vi→zh, zh→mn, mn→zh, zh→cs and cs→zh translation respectively, showing the effectiveness of our approach.

II. Related Work

A. Low-resource Translation

To address the over-fitting problem of NMT, [37] propose to jointly train a forward translation model and a reverse model. [5], [45] reduce the number of parameters by reducing the depth, hidden dimension and sub-word vocabulary, bind embedding matrices with the classifier weight matrix, and use a stronger regularization strategy. [19] use shallow models to efficiently search hyper-parameters of deep models on the CCMT 2022 Chinese-Thai low-resource translation task.

To introduce additional knowledge for low-resource scenarios, [36] fine-tune the model trained on high-resource translation tasks on the low-resource task. [25] investigate style transfer for languages without style-labelled corpora. [61] leverage embedding duplication between aligned sub-words to extend the Parent-Child transfer learning method. [30] propose to continuously transfer knowledge from the parent model during the training of the child model. [20] enlarge the dataset through paraphrasing source sentences.

Corresponding author: Hongfei Xu. This work is partially supported by the National Natural Science Foundation of China (Grant No. 62306284), the Natural Science Foundation of Henan Province (Grant No. 232300421386), and the Henan Provincial Science and Technology Research Project (Grant No. 232102211041).

979-8-3503-3078-6/23/$31.00 ©2023 IEEE

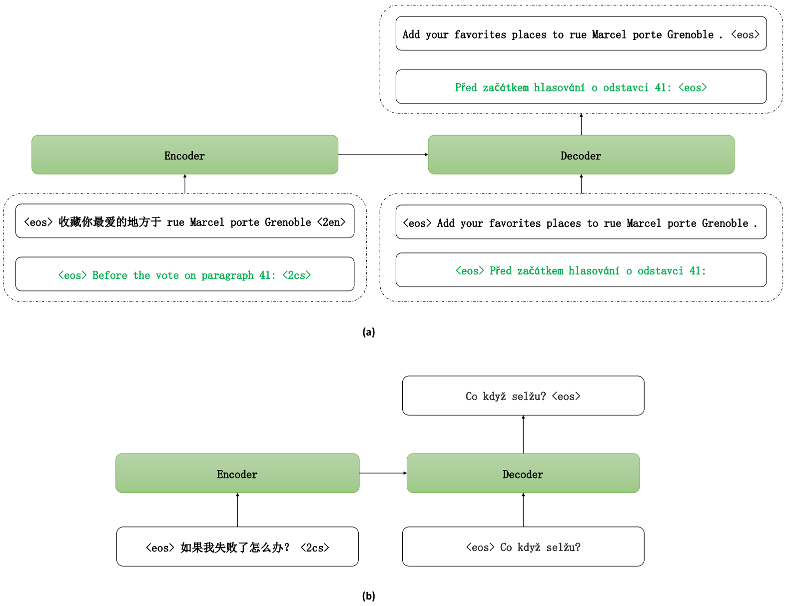

Fig. 1. Using English-centric pivoted parallel data with MNMT and two-stage fine-tuning. (a): fine-tune the model on the zh→En task and the En→X task, (b): fine-tune the model from the first stage on the zh→X translation task.

B. Multilingual Neural Machine Translation

Multilingual NMT includes one-to-many [13], many-to-many [15] and zero-shot [16] scenarios. A simple solution is to insert a target language token at the beginning of the input sentence [22].

Multilingual NMT has to handle different languages in one joint representation space, neglecting their linguistic diversity, especially for massively multilingual NMT [1], [3], [17], [21], [35], [63]. Most studies focus on how to mitigate this representation bottleneck [7], [9], [34], [40], [50], [51], [54], [56], [57], [60], [66], [67].

There are also studies on the trade-off between shared and language-specific parameters [43], [62], on the training of multilingual NMT [4], [32], [46], [55], [58], on adapting MNMT models to multiple domains [8], [27] and on analyzing translations from multilingual NMT [28], [65] or the trained model [10], [14], [18], [26], [38], [48], Transferring a pre-trained multilingual NMT model can improve the performance of downstream language pairs [23], [31], [48], especially for Chinese-centric Low-resource scenarios [11]. Multilingual data also has been proven useful for unsupervised NMT [44], [47].

III. Leveraging English-centric Pivoted Parallel Data for Chinese-centric Translation via MNMT

A. Utilizing English-centric Parallel Data via MNMT

Multilingual translation between multiple language pairs with a single model [1], [15], [22] enables transfer learning across languages and zero-shot translation. Thus, for the translation between Chinese and other languages zh→X, we can model the translation tasks zh→X, zh→En and En→X in a single model in the MNMT manner to transfer the alignment and translation knowledge of the English-centric parallel data to the translation task from zh to X.

Specifically, we replace the special start-of-sentence token with the target language token at both the source and the target side to indicate the translation direction, and train the model on the concatenation of the training data of involved translation tasks. We use the target language token instead of the source language token at the source side, because [59] find that translation already happens in the encoder layers, and providing the target language token to the encoder can help it figure out the target language. This is also the common practice of many MNMT studies, like OPUS-100 [64].

a) Data Sampling for MNMT: There can be huge differences in the amount of parallel data of different translation tasks. When training on the data of more than one translation task, we have to balance the dataset based on the importance of the task by sampling.

Assume the number of training data of the translation task t is D_t, and the original proportion p_t of the task is shown in Equation 1.

$$p_t = \frac{D_t}{\sum D_i} \tag{1}$$

Following [2], we introduce a factor T as the sampling temperature to balance between involved tasks, and the sampling probability p_t^s of task t is defined in Equation 2.

$$p_t^s = p_t^{\frac{1}{T}} \tag{2}$$

where $T = 1$ follows the original data distribution and the sampling distribution is close to uniform using a large value.

TABLE I
MAIN RESULTS.

Translation task		zh→vi	vi→zh	zh→mn	mn→zh	zh→cs	cs→zh
#Sentece pairs	zh↔en	30499553					
	X↔en	1004000		4294		107314898	
	zh↔X	197995		197993		197995	
Baseline		51.63	48.47	39.76	34.73	28.55	46.98
Ours		**53.28**	**48.71**	**40.67**	**38.20**	**31.43**	**53.33**

As the A→En and the En→B task are equivalently important for the translation from A to B, we use infinite ($T = \infty$) by default in practice.

b) Vocabulary Pruning: The mBART pre-trained model is very large, covering 25 languages, and has a vocabulary size of around 250k tokens. The large vocabulary brings challenges to its fine-tuning, on both GPU memory and computing efficiency. To address this, we first collect the tokens in the parallel data of the translation tasks, and then pruning out the other tokens that do not appear in the parallel data from the vocabulary (and their corresponding source embeddings, target embeddings and classifier weights and biases) of mBART. In practice, this can effectively reduce the around half of the vocabulary size, enabling faster mBART fine-tuning with larger batch sizes.

B. Two-stage Fine-tuning

It is a problem to balance between the translation task zh→X and the translation tasks zh→En and En→X if we train the MNMT model in a single stage, as zh→X is our target task which shall be assigned with a higher priority than the others during sampling, but over-sampling the training data for the zh→X translation task on the other hand hinders the model to learn from the zh→En task and the En→X task.

To address this issue, we employ a 2-stage fine-tuning method, where we first fine-tune the model on the zh→En task and the En→X task, and then fine-tune the model from the first stage on the zh→X translation task, as shown in Figure 1. In this way, the first fine-tuning focuses on learning translation and alignment knowledge from the English-centric tasks, and the second stage fully concentrates on the zh→X translation task while utilizing the translation and alignment knowledge gained from the English-centric parallel data in the first stage.

IV. EXPERIMENTS

A. Settings

a) Datasets: We tested the effectiveness of our approach on the CCMT 2023 low-resource translation task between Chinese and the languages among "the Belt and Road", specifically Vietnamese (vi), Mongolian (mn) and Czech (cs). As the shared task only provides the training sets for now, we held out the last 2000 sentence pairs of the training set as the validation set for evaluation for the ease of reproduction.

For the English-centric parallel data, we used the corresponding sections of OPUS-100 [1], [52], [64] for Vietnamese and Mongolian, part of the training data of the WMT 2022 news translation task [24] for Czech (except for back-translated data) and Chinese (ParaCrawl, News Commentary and UN Parallel Corpus). We use the mBART tokenizer for cs, vi, mn. We only used sequences with no more than 512 tokens for training.

TABLE II
ZERO-SHOT TRANSLATION PERFORMANCE.

zh→vi	vi→zh	zh→mn	mn→zh	zh→cs	cs→zh
8.19	26.29	0.07	10.57	10.10	15.64

TABLE III
DATA BALANCING RESULTS.

T	zh→vi	vi→zh	zh→mn	mn→zh	zh→cs	cs→zh
1	53.13	48.31	40.61	37.79	31.37	49.53
∞	**53.28**	**48.71**	**40.67**	**38.20**	**31.43**	**53.33**

b) Model Settings: we adopted the large BART models with 12 encoder and decoder layers, 1024 as the embedding dimension and 4096 as the hidden dimension for fine-tuning. The residual connection dropout probability was 0.1, and the activation dropout and attention dropout were disabled. We used the Adam optimizer and a learning rate of $1e-5$ for fine-tuning. We used mBART25 for Czech and Vietnamese, and mBART50 for Mongolian.

c) Evaluation: We decoded with a beam size of 4, and evaluated the translation quality by character-level BLEU [39] with the SacreBLEU toolkit [41].

B. Main Results

Our main results are shown Table I. Table I shows that: 1) our approach can lead to consistent and significant improvements ($+1.65$, $+0.24$, $+0.91$, $+3.47$, $+2.88$, $+6.35$ BLEU for zh→vi, vi→zh, zh→mn, mn→zh, zh→cs and cs→zh respectively) on all evaluated tasks. 2) the cs↔zh has the largest amount of English-centric parallel data and leads to the largest improvements.

C. Zero-shot Evaluation

MNMT enables not only transfer learning but also zero-shot translation. The model trained on the zh→en and en→X tasks from the first stage shall already have some abilities in zh→X translation. We tested the performance of the first stage model on the validation sets, and results are shown in Table II.

Table II verifies that the model already gains some translation and alignment knowledge for the zh→X task from the parallel data of zh→en and en→X tasks. The poor performance of the zh↔mn task may due to the lack of data (Table I).

D. Data Balancing

We study the effects of the sampling temperature T on performance, and results are shown in Table III.

Table III shows that using uniform sampling in the first stage ($T = \infty$) consistently leads to better performance than sampling according to the original distribution.

V. CONCLUSION

In this paper, we propose to utilize English-centric parallel data to improve the performance of low-resource translation tasks between Chinese and other languages by MNMT modeling and two-stage fine-tuning. Experiments on the CCMT 2023 low-resource translation tasks between Chinese and the languages among "the Belt and Road" shows that our approach can consistently lead to significant improvements (+1.65, +0.24, +0.91, +3.47, +2.88, +6.35 BLEU for zh→vi, vi→zh, zh→mn, mn→zh, zh→cs and cs→zh respectively) over the strong mBART baseline.

REFERENCES

[1] Aharoni, R., Johnson, M., Firat, O.: Massively multilingual neural machine translation. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 3874–3884. Association for Computational Linguistics, Minneapolis, Minnesota (Jun 2019). https://doi.org/10.18653/v1/N19-1388, https://www.aclweb.org/anthology/N19-1388

[2] Aharoni, R., Johnson, M., Firat, O.: Massively multilingual neural machine translation. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 3874–3884. Association for Computational Linguistics, Minneapolis, Minnesota (Jun 2019). https://doi.org/10.18653/v1/N19-1388, https://aclanthology.org/N19-1388

[3] Ahuja, K., Sitaram, S., Dandapat, S., Choudhury, M.: On the calibration of massively multilingual language models. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. pp. 4310–4323. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022), https://aclanthology.org/2022.emnlp-main.290

[4] Al-Shedivat, M., Parikh, A.: Consistency by agreement in zero-shot neural machine translation. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 1184–1197. Association for Computational Linguistics, Minneapolis, Minnesota (Jun 2019). https://doi.org/10.18653/v1/N19-1121, https://www.aclweb.org/anthology/N19-1121

[5] Araabi, A., Monz, C.: Optimizing transformer for low-resource neural machine translation. In: Proceedings of the 28th International Conference on Computational Linguistics. pp. 3429–3435. International Committee on Computational Linguistics, Barcelona, Spain (Online) (Dec 2020). https://doi.org/10.18653/v1/2020.coling-main.304, https://aclanthology.org/2020.coling-main.304

[6] Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. CoRR **abs/1409.0473** (2014)

[7] Bapna, A., Firat, O.: Simple, scalable adaptation 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). pp. 1538–1548. Association for Computational Linguistics, Hong Kong, China (Nov 2019). https://doi.org/10.18653/v1/D19-1165, https://www.aclweb.org/anthology/D19-1165

[8] Baziotis, C., Artetxe, M., Cross, J., Bhosale, S.: Multilingual machine translation with hyper-adapters. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. pp. 1170–1185. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022), https://aclanthology.org/2022.emnlp-main.77

[9] Blackwood, G., Ballesteros, M., Ward, T.: Multilingual neural machine translation with task-specific attention. In: Proceedings of the 27th International Conference on Computational Linguistics. pp. 3112–3122. Association for Computational Linguistics, Santa Fe, New Mexico, USA (Aug 2018), https://www.aclweb.org/anthology/C18-1263

[10] Cheng, Y., Bapna, A., Firat, O., Cao, Y., Wang, P., Macherey, W.: Multilingual mix: Example interpolation improves multilingual neural machine translation. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 4092–4102. Association for Computational Linguistics, Dublin, Ireland (May 2022). https://doi.org/10.18653/v1/2022.acl-long.282, https://aclanthology.org/2022.acl-long.282

[11] Dabre, R., Fujita, A., Chu, C.: Exploiting multilingualism through multistage fine-tuning for low-resource 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). pp. 1410–1416. Association for Computational Linguistics, Hong Kong, China (Nov 2019). https://doi.org/10.18653/v1/D19-1146, https://www.aclweb.org/anthology/D19-1146

[12] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota (Jun 2019). https://doi.org/10.18653/v1/N19-1423, https://aclanthology.org/N19-1423

[13] Dong, D., Wu, H., He, W., Yu, D., Wang, H.: 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 (Volume 1: Long Papers). pp. 1723–1732. Association for Computational Linguistics, Beijing, China (Jul 2015). https://doi.org/10.3115/v1/P15-1166, https://www.aclweb.org/anthology/P15-1166

[14] Eriguchi, A., Xie, S., Qin, T., Hassan, H.: Building multilingual machine translation systems that serve arbitrary XY translations. In: Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 600–606. Association for Computational Linguistics, Seattle, United States (Jul 2022). https://doi.org/10.18653/v1/2022.naacl-main.44, https://aclanthology.org/2022.naacl-main.44

[15] Firat, O., Cho, K., Bengio, Y.: Multi-way, multilingual neural machine translation with a shared attention mechanism. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 866–875. Association for Computational Linguistics, San Diego, California (Jun 2016). https://doi.org/10.18653/v1/N16-1101, https://www.aclweb.org/anthology/N16-1101

[16] Firat, O., Sankaran, B., Al-onaizan, Y., Yarman Vural, F.T., Cho, K.: Zero-resource translation with multi-lingual neural machine translation. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. pp. 268–277. Association for Computational Linguistics, Austin, Texas (Nov 2016). https://doi.org/10.18653/v1/D16-1026, https://www.aclweb.org/anthology/D16-1026

[17] Freitag, M., Firat, O.: Complete multilingual neural machine translation. In: Proceedings of the Fifth Conference on Machine Translation. pp. 550–560. Association for Computational Linguistics, Online (Nov 2020), https://www.aclweb.org/anthology/2020.wmt-1.66

[18] Gu, S., Feng, Y.: Improving zero-shot multilingual translation with universal representations and cross-mapping. In: Findings of the Association for Computational Linguistics: EMNLP 2022. pp. 6492–6504. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022), https://aclanthology.org/2022.findings-emnlp.485

[19] Hao, W., Xu, H., Mu, L., Zan, H.: Optimizing deep transformers for chinese-thai low-resource translation. In: Xiao, T., Pino, J. (eds.) Machine Translation. pp. 117–126. Springer Nature Singapore, Singapore (2022)

[20] Hao, W., Xu, H., Xiong, D., Zan, H., Mu, L.: ParaZh-22M: A large-scale Chinese parabank via machine translation. In: Proceedings of the 29th International Conference on Computational Linguistics. pp. 3885–3897. International Committee on Computational Linguistics, Gyeongju, Republic of Korea (Oct 2022), `https://aclanthology.org/2022.coling-1.341`

[21] Huang, Y., Feng, X., Geng, X., Qin, B.: Unifying the convergences in multilingual neural machine translation. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. pp. 6822–6835. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022), `https://aclanthology.org/2022.emnlp-main.458`

[22] Johnson, M., Schuster, M., Le, Q.V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G., Hughes, M., Dean, J.: Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. arXiv e-prints arXiv:1611.04558 (Nov 2016). https://doi.org/10.48550/arXiv.1611.04558

[23] Kim, Y., Gao, Y., Ney, H.: Effective cross-lingual transfer of neural machine translation models without shared vocabularies. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 1246–1257. Association for Computational Linguistics, Florence, Italy (Jul 2019). https://doi.org/10.18653/v1/P19-1120, `https://www.aclweb.org/anthology/P19-1120`

[24] Kocmi, T., Bawden, R., Bojar, O., Dvorkovich, A., Federmann, C., Fishel, M., Gowda, T., Graham, Y., Grundkiewicz, R., Haddow, B., Knowles, R., Koehn, P., Monz, C., Morishita, M., Nagata, M., Nakazawa, T., Novák, M., Popel, M., Popović, M., Shmatova, M.: Findings of the 2022 conference on machine translation (wmt22). In: Proceedings of the Seventh Conference on Machine Translation. pp. 1–45. Association for Computational Linguistics, Abu Dhabi (December 2022), `https://aclanthology.org/2022.wmt-1.1`

[25] Krishna, K., Nathani, D., Garcia, X., Samanta, B., Talukdar, P.: Few-shot controllable style transfer for low-resource multilingual settings. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 7439–7468. Association for Computational Linguistics, Dublin, Ireland (May 2022). https://doi.org/10.18653/v1/2022.acl-long.514, `https://aclanthology.org/2022.acl-long.514`

[26] Kudugunta, S., Bapna, A., Caswell, I., Firat, O.: Investigating multilingual NMT representations at scale. 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). pp. 1565–1575. Association for Computational Linguistics, Hong Kong, China (Nov 2019). https://doi.org/10.18653/v1/D19-1167, `https://www.aclweb.org/anthology/D19-1167`

[27] Lai, W., Chronopoulou, A., Fraser, A.: m^4 adapter: Multilingual multi-domain adaptation for machine translation with a meta-adapter. In: Findings of the Association for Computational Linguistics: EMNLP 2022. pp. 4282–4296. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022), `https://aclanthology.org/2022.findings-emnlp.315`

[28] Lakew, S.M., Cettolo, M., Federico, M.: A comparison of transformer and recurrent neural networks on multilingual neural machine translation. In: Proceedings of the 27th International Conference on Computational Linguistics. pp. 641–652. Association for Computational Linguistics, Santa Fe, New Mexico, USA (Aug 2018), `https://www.aclweb.org/anthology/C18-1054`

[29] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L.: BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 7871–7880. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.703, `https://aclanthology.org/2020.acl-main.703`

[30] Li, Z., Liu, X., Wong, D.F., Chao, L.S., Zhang, M.: ConsistTL: Modeling consistency in transfer learning for low-resource neural machine translation. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. pp. 8383–8394. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022), `https://aclanthology.org/2022.emnlp-main.574`

[31] Lin, Z., Pan, X., Wang, M., Qiu, X., Feng, J., Zhou, H., Li, L.: Pre-training multilingual neural machine translation by leveraging alignment information. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 2649–2663. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.emnlp-main.210, `https://www.aclweb.org/anthology/2020.emnlp-main.210`

[32] Liu, J., Huang, K., Li, J., Liu, H., Su, J., Huang, D.: Adaptive token-level cross-lingual feature mixing for multilingual neural machine translation. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. pp. 10097–10113. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022), `https://aclanthology.org/2022.emnlp-main.687`

[33] Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M., Zettlemoyer, L.: Multilingual Denoising Pre-training for Neural Machine Translation. arXiv e-prints arXiv:2001.08210 (Jan 2020). https://doi.org/10.48550/arXiv.2001.08210

[34] Lyu, S., Son, B., Yang, K., Bae, J.: Revisiting Modularized Multilingual NMT to Meet Industrial Demands. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 5905–5918. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.emnlp-main.476, `https://www.aclweb.org/anthology/2020.emnlp-main.476`

[35] Mohammadshahi, A., Nikoulina, V., Berard, A., Brun, C., Henderson, J., Besacier, L.: SMaLL-100: Introducing shallow multilingual machine translation model for low-resource languages. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. pp. 8348–8359. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022), `https://aclanthology.org/2022.emnlp-main.571`

[36] Nguyen, T.Q., Chiang, D.: 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). pp. 296–301. Asian Federation of Natural Language Processing, Taipei, Taiwan (Nov 2017), `https://aclanthology.org/I17-2050`

[37] Niu, X., Xu, W., Carpuat, M.: Bi-directional differentiable input reconstruction for low-resource neural machine translation. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 442–448. Association for Computational Linguistics, Minneapolis, Minnesota (Jun 2019). https://doi.org/10.18653/v1/N19-1043, `https://aclanthology.org/N19-1043`

[38] Oncevay, A., Haddow, B., Birch, A.: Bridging linguistic typology and multilingual machine translation with multi-view language representations. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 2391–2406. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.emnlp-main.187, `https://www.aclweb.org/anthology/2020.emnlp-main.187`

[39] Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. pp. 311–318. Association for Computational Linguistics, Philadelphia, Pennsylvania, USA (Jul 2002). https://doi.org/10.3115/1073083.1073135, `https://www.aclweb.org/anthology/P02-1040`

[40] Platanios, E.A., Sachan, M., Neubig, G., Mitchell, T.: Contextual parameter generation for universal neural machine translation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. pp. 425–435. Association for Computational Linguistics, Brussels, Belgium (Oct-Nov 2018). https://doi.org/10.18653/v1/D18-1039, `https://www.aclweb.org/anthology/D18-1039`

[41] Post, M.: A call for clarity in reporting BLEU scores. In: Proceedings of the Third Conference on Machine Translation: Research Papers. pp. 186–191. Association for Computational Linguistics, Brussels, Belgium (Oct 2018). https://doi.org/10.18653/v1/W18-6319, `https://www.aclweb.org/anthology/W18-6319`

[42] Raffel, C., Shazeer, N.M., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J.: Exploring the limits of transfer learning with a unified text-to-text transformer. ArXiv **abs/1910.10683** (2019)

[43] Sachan, D., Neubig, G.: Parameter sharing methods for multilingual self-attentional translation models. In: Proceedings of the Third Conference on Machine Translation: Research Papers. pp. 261–271. Association for Computational Linguistics, Brussels, Belgium (Oct 2018). https://doi.org/10.18653/v1/W18-6327, https://www.aclweb.org/anthology/W18-6327

[44] Sen, S., Gupta, K.K., Ekbal, A., Bhattacharyya, P.: Multilingual unsupervised NMT using shared encoder and language-specific decoders. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 3083–3089. Association for Computational Linguistics, Florence, Italy (Jul 2019). https://doi.org/10.18653/v1/P19-1297, https://www.aclweb.org/anthology/P19-1297

[45] Sennrich, R., Zhang, B.: Revisiting low-resource neural machine translation: A case study. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 211–221. Association for Computational Linguistics, Florence, Italy (Jul 2019). https://doi.org/10.18653/v1/P19-1021, https://aclanthology.org/P19-1021

[46] Siddhant, A., Bapna, A., Cao, Y., Firat, O., Chen, M., Kudugunta, S., Arivazhagan, N., Wu, Y.: Leveraging monolingual data with self-supervision for multilingual neural machine translation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 2827–2835. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.252, https://www.aclweb.org/anthology/2020.acl-main.252

[47] Sun, H., Wang, R., Chen, K., Utiyama, M., Sumita, E., Zhao, T.: Knowledge distillation for multilingual unsupervised neural machine translation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 3525–3535. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.324, https://www.aclweb.org/anthology/2020.acl-main.324

[48] Sun, S., Fan, A., Cross, J., Chaudhary, V., Tran, C., Koehn, P., Guzmán, F.: Alternative input signals ease transfer in multilingual machine translation. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 5291–5305. Association for Computational Linguistics, Dublin, Ireland (May 2022). https://doi.org/10.18653/v1/2022.acl-long.363, https://aclanthology.org/2022.acl-long.363

[49] Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. ArXiv **abs/1409.3215** (2014)

[50] Tan, X., Chen, J., He, D., Xia, Y., Qin, T., Liu, T.Y.: 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). pp. 963–973. Association for Computational Linguistics, Hong Kong, China (Nov 2019). https://doi.org/10.18653/v1/D19-1089, https://www.aclweb.org/anthology/D19-1089

[51] Tan, X., Ren, Y., He, D., Qin, T., Liu, T.Y.: Multilingual neural machine translation with knowledge distillation. In: International Conference on Learning Representations (2019), https://openreview.net/forum?id=S1gUsoR9YX

[52] Tiedemann, J.: Parallel data, tools and interfaces in opus. In: Language Resources and Evaluation (2012)

[53] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention Is All You Need. arXiv e-prints arXiv:1706.03762 (Jun 2017). https://doi.org/10.48550/arXiv.1706.03762

[54] Wang, X., Pham, H., Arthur, P., Neubig, G.: Multilingual neural machine translation with soft decoupled encoding. In: International Conference on Learning Representations (2019), https://openreview.net/forum?id=Skeke3C5Fm

[55] Wang, X., Tsvetkov, Y., Neubig, G.: Balancing training for multilingual neural machine translation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 8526–8537. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.754, https://www.aclweb.org/anthology/2020.acl-main.754

[56] Wang, Y., Zhang, J., Zhai, F., Xu, J., Zong, C.: Three strategies to improve one-to-many multilingual translation. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. pp. 2955–2960. Association for Computational Linguistics, Brussels, Belgium (Oct-Nov 2018). https://doi.org/10.18653/v1/D18-1326, https://www.aclweb.org/anthology/D18-1326

[57] Wang, Y., Zhou, L., Zhang, J., Zhai, F., Xu, J., Zong, C.: A compact and language-sensitive multilingual translation method. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 1213–1223. Association for Computational Linguistics, Florence, Italy (Jul 2019). https://doi.org/10.18653/v1/P19-1117, https://www.aclweb.org/anthology/P19-1117

[58] Wang, Y., Zhai, C., Hassan, H.: Multi-task learning for multilingual neural machine translation. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1022–1034. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.emnlp-main.75, https://www.aclweb.org/anthology/2020.emnlp-main.75

[59] Xu, H., van Genabith, J., Liu, Q., Xiong, D.: Probing word translations in the transformer and trading decoder for encoder layers. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 74–85. Association for Computational Linguistics, Online (Jun 2021). https://doi.org/10.18653/v1/2021.naacl-main.7, https://aclanthology.org/2021.naacl-main.7

[60] Xu, H., Liu, Q., van Genabith, J., Xiong, D., Zhang, M.: Multi-head highly parallelized LSTM decoder for neural machine translation. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). pp. 273–282. Association for Computational Linguistics, Online (Aug 2021). https://doi.org/10.18653/v1/2021.acl-long.23, https://aclanthology.org/2021.acl-long.23

[61] Xu, M., Hong, Y.: Sub-word alignment is still useful: A vest-pocket method for enhancing low-resource machine translation. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 613–619. Association for Computational Linguistics, Dublin, Ireland (May 2022). https://doi.org/10.18653/v1/2022.acl-short.68, https://aclanthology.org/2022.acl-short.68

[62] Zhang, B., Bapna, A., Sennrich, R., Firat, O.: Share or not? learning to schedule language-specific capacity for multilingual translation. In: International Conference on Learning Representations (2021), https://openreview.net/forum?id=Wj4ODo0uyCF

[63] Zhang, B., Williams, P., Titov, I., Sennrich, R.: Improving massively multilingual neural machine translation and zero-shot translation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 1628–1639. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.148, https://www.aclweb.org/anthology/2020.acl-main.148

[64] Zhang, B., Williams, P., Titov, I., Sennrich, R.: Improving massively multilingual neural machine translation and zero-shot translation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 1628–1639. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.148, https://aclanthology.org/2020.acl-main.148

[65] Zhang, H., Zhang, S., Xiang, Y., Liang, B., Su, J., Miao, Z., Wang, H., Xu, R.: CLLE: A benchmark for continual language learning evaluation in multilingual machine translation. In: Findings of the Association for Computational Linguistics: EMNLP 2022. pp. 428–443. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Dec 2022), https://aclanthology.org/2022.findings-emnlp.30

[66] Zhu, C., Yu, H., Cheng, S., Luo, W.: Language-aware interlingua for multilingual neural machine translation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 1650–1655. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.150, https://www.aclweb.org/anthology/2020.acl-main.150

[67] Zoph, B., Knight, K.: Multi-source neural translation. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 30–34. Association for Computational Linguistics, San Diego, California (Jun 2016). https://doi.org/10.18653/v1/N16-1004, https://www.aclweb.org/anthology/N16-1004