

REFERENCES

- [1] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens *et al.*, “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*, 2007, pp. 177–180.
- [2] A. Lopez, “Statistical machine translation,” *ACM Computing Surveys (CSUR)*, vol. 40, no. 3, pp. 1–49, 2008.
- [3] S. Ranathunga, E.-S. A. Lee, M. Prifti Skenduli, R. Shekhar, M. Alam, and R. Kaur, “Neural machine translation for low-resource languages: A survey,” *ACM Computing Surveys*, vol. 55, no. 11, pp. 1–37, 2023.
- [4] R. Dabre, C. Chu, and A. Kunchukuttan, “A survey of multilingual neural machine translation,” *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–38, 2020.
- [5] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” *arXiv preprint arXiv:1409.0473*, 2014.
- [6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [7] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” *arXiv preprint arXiv:1409.3215*, 2014.
- [8] R. Sennrich and B. Zhang, “Revisiting low-resource neural machine translation: A case study,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, A. Korhonen, D. Traum, and L. Màrquez, Eds. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 211–221. [Online]. Available: <https://aclanthology.org/P19-1021/>
- [9] S. Ranathunga, E.-S. A. Lee, M. P. Skenduli, R. Shekhar, M. Alam, and R. Kaur, “Neural machine translation for low-resource languages: A survey,” 2021.
- [10] B. Haddow, R. Bawden, A. V. M. Barone, J. Helcl, and A. Birch, “Survey of low-resource machine translation,” *Computational Linguistics*, vol. 48, no. 3, pp. 673–732, 2022.
- [11] R. Östling and J. Tiedemann, “Neural machine translation for low-resource languages,” *arXiv preprint arXiv:1708.05729*, 2017.

- [12] S. Manchanda and G. Grunin, “Domain informed neural machine translation: Developing translation services for healthcare enterprise,” in *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, A. Martins, H. Moniz, S. Fumega, B. Martins, F. Batista, L. Coheur, C. Parra, I. Trancoso, M. Turchi, A. Bisazza, J. Moorkens, A. Guerbero, M. Nurminen, L. Marg, and M. L. Forcada, Eds. Lisboa, Portugal: European Association for Machine Translation, Nov. 2020, pp. 255–261. [Online]. Available: <https://aclanthology.org/2020.eamt-1.27/>
- [13] D. Saunders, “Domain adaptation and multi-domain adaptation for neural machine translation: A survey,” *Journal of Artificial Intelligence Research*, vol. 75, pp. 351–424, 2022.
- [14] P. Koehn and R. Knowles, “Six challenges for neural machine translation,” in *Proceedings of the First Workshop on Neural Machine Translation*, T. Luong, A. Birch, G. Neubig, and A. Finch, Eds. Vancouver: Association for Computational Linguistics, Aug. 2017, pp. 28–39. [Online]. Available: <https://aclanthology.org/W17-3204/>
- [15] C. Chu and R. Wang, “A survey of domain adaptation for neural machine translation,” in *Proceedings of the 27th International Conference on Computational Linguistics*, E. M. Bender, L. Derczynski, and P. Isabelle, Eds. Santa Fe, New Mexico, USA: Association for Computational Linguistics, Aug. 2018, pp. 1304–1319. [Online]. Available: <https://aclanthology.org/C18-1111/>
- [16] A. Currey, P. Mathur, and G. Dinu, “Distilling multiple domains for neural machine translation,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds. Online: Association for Computational Linguistics, Nov. 2020, pp. 4500–4511. [Online]. Available: <https://aclanthology.org/2020.emnlp-main.364/>
- [17] Y. Liang, F. Meng, J. Wang, J. Xu, Y. Chen, and J. Zhou, “Continual learning with semi-supervised contrastive distillation for incremental neural machine translation,” in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, L.-W. Ku, A. Martins, and V. Srikumar, Eds. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 10914–10928. [Online]. Available: <https://aclanthology.org/2024.acl-long.588/>
- [18] X. Pan, M. Wang, L. Wu, and L. Li, “Contrastive learning for many-to-many multilingual 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)*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds. Online: Association for Computational Linguistics, Aug. 2021, pp. 244–258. [Online]. Available: <https://aclanthology.org/2021.acl-long.21/>
- [19] M. Pham, J. Crego, F. Yvon, and J. Senellart, “Generic and specialized word embeddings for multi-domain machine translation,” in *Proceedings of the 16th International Conference on Spoken Language Translation*, J. Niehues, R. Cattoni, S. Stüker, M. Negri, M. Turchi, T.-L. Ha, E. Salesky, R. Sanabria, L. Barrault, L. Specia, and M. Federico, Eds. Hong Kong: Association for Computational Linguistics, Nov. 2-3 2019. [Online]. Available: <https://aclanthology.org/2019.iwslt-1.26/>
- [20] G. Hinton, “Distilling the knowledge in a neural network,” *arXiv preprint arXiv:1503.02531*, 2015.
- [21] Y. Kim and A. M. Rush, “Sequence-level knowledge distillation,” in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, J. Su, K. Duh, and X. Carreras, Eds. Austin, Texas: Association for Computational Linguistics, Nov. 2016, pp. 1317–1327. [Online]. Available: <https://aclanthology.org/D16-1139/>
- [22] C. Chu, R. Dabre, and S. Kurohashi, “An empirical comparison of domain adaptation methods for neural machine translation,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, R. Barzilay and M.-Y. Kan, Eds. Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 385–391. [Online]. Available: <https://aclanthology.org/P17-2061/>
- [23] R. Dabre, A. Fujita, and C. Chu, “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)*, K. Inui, J. Jiang, V. Ng, and X. Wan, Eds. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 1410–1416. [Online]. Available: <https://aclanthology.org/D19-1146/>
- [24] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” 2016. [Online]. Available: <https://arxiv.org/abs/1409.0473>

- [25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- [26] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734. [Online]. Available: <https://aclanthology.org/D14-1179/>
- [27] B. Ghojogh and A. Ghodsi, “Recurrent neural networks and long short-term memory networks: Tutorial and survey,” 2023. [Online]. Available: <https://arxiv.org/abs/2304.11461>
- [28] R. C. Staudemeyer and E. R. Morris, “Understanding lstm – a tutorial into long short-term memory recurrent neural networks,” 2019. [Online]. Available: <https://arxiv.org/abs/1909.09586>
- [29] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” 2014. [Online]. Available: <https://arxiv.org/abs/1412.3555>
- [30] S. A. Mohamed, A. A. Elsayed, Y. Hassan, and M. A. Abdou, “Neural machine translation: past, present, and future,” *Neural Computing and Applications*, vol. 33, pp. 15 919–15 931, 2021.
- [31] M. X. Chen, O. Firat, A. Bapna, M. Johnson, W. Macherey, G. Foster, L. Jones, N. Parmar, M. Schuster, Z. Chen, Y. Wu, and M. Hughes, “The best of both worlds: Combining recent advances in neural machine translation,” 2018. [Online]. Available: <https://arxiv.org/abs/1804.09849>
- [32] M. Ott, S. Edunov, D. Grangier, and M. Auli, “Scaling neural machine translation,” 2018. [Online]. Available: <https://arxiv.org/abs/1806.00187>
- [33] M. Dehghani, S. Gouws, O. Vinyals, J. Uszkoreit, and Łukasz Kaiser, “Universal transformers,” 2019. [Online]. Available: <https://arxiv.org/abs/1807.03819>

- [34] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, “Albert: A lite bert for self-supervised learning of language representations,” 2020. [Online]. Available: <https://arxiv.org/abs/1909.11942>
- [35] P. Michel, O. Levy, and G. Neubig, “Are sixteen heads really better than one?” 2019. [Online]. Available: <https://arxiv.org/abs/1905.10650>
- [36] J. Kasai, H. Peng, Y. Zhang, D. Yogatama, G. Ilharco, N. Pappas, Y. Mao, W. Chen, and N. A. Smith, “Finetuning pretrained transformers into RNNs,” in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, Eds. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 10 630–10 643. [Online]. Available: <https://aclanthology.org/2021.emnlp-main.830/>
- [37] N. Team, M. R. Costa-jussà, J. Cross, O. Çelebi, M. Elbayad, K. Heafield, K. Heffernan, E. Kalbassi, J. Lam, D. Licht, J. Maillard, A. Sun, S. Wang, G. Wenzek, A. Youngblood, B. Akula, L. Barrault, G. M. Gonzalez, P. Hansanti, ..., and J. Wang, “No Language Left Behind: Scaling Human-Centered Machine Translation,” 2022. [Online]. Available: <https://arxiv.org/abs/2207.04672>
- [38] A. Fan, S. Bhosale, H. Schwenk, Z. Ma, A. El-Kishky, S. Goyal, M. Baines, O. Celebi, G. Wenzek, V. Chaudhary, N. Goyal, T. Birch, V. Liptchinsky, S. Edunov, E. Grave, M. Auli, and A. Joulin, “Beyond english-centric multilingual machine translation,” 2020. [Online]. Available: <https://arxiv.org/abs/2010.11125>
- [39] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, and C. Raffel, “mT5: A massively multilingual pre-trained text-to-text transformer,” in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, and Y. Zhou, Eds. Online: Association for Computational Linguistics, Jun. 2021, pp. 483–498. [Online]. Available: <https://aclanthology.org/2021.naacl-main.41/>
- [40] Y. Liu, J. Gu, N. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis, and L. Zettlemoyer, “Multilingual denoising pre-training for neural machine translation,” *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 726–742, 2020. [Online]. Available: <https://aclanthology.org/2020.tacl-1.47/>

- [41] H. M. Alemayehu, H. M. Zahera, and A.-C. Ngonga Ngomo, “Error analysis of multilingual language models in machine translation: A case study of English-Amharic translation,” in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 19 758–19 768. [Online]. Available: <https://aclanthology.org/2024.emnlp-main.1102/>
- [42] E.-S. A. Lee, S. Thillainathan, S. Nayak, S. Ranathunga, D. I. Adelani, R. Su, and A. D. McCarthy, “Pre-trained multilingual sequence-to-sequence models: A hope for low-resource language translation?” in *Findings of the Association for Computational Linguistics: ACL 2022*, S. Muresan, P. Nakov, and A. Villavicencio, Eds. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 58–67. [Online]. Available: <https://aclanthology.org/2022.findings-acl.6/>
- [43] A. El-Kishky, V. Chaudhary, A. Fan, S. Bhosale, H. Schwenk, Zhiyi, S. Goyal, M. Baines, O. Çelebi, G. Wenzek, A. Fernando, S. Ranathunga, N. Goyal, C. Gao, J. Chen, D. Ju, S. Kr-373, M. Ranzato, F. Guzmán, F. Guzmán, P.-J. Chen, M. Ott, G. Pino, P. Lample, and V. Koehn, “Sequence-to-Sequence Multilingual Pre-Trained Models: A Hope for Low-Resource Language Translation?” 2021. [Online]. Available: <https://api.semanticscholar.org/CorpusID:252620124>
- [44] C. Cieri, M. Maxwell, S. Strassel, J. Devlin, M.-W. Chang, K. Lee, D. M. Eberhard, G. F. Simons, K. Kann, O. Lacroix, and A. Sørby, “Exploring the low-resource transfer-learning with mt5 model,” 2022. [Online]. Available: <https://api.semanticscholar.org/CorpusID:252547526>
- [45] T. J. Singh, S. R. Singh, and P. Sarmah, “Can big models help diverse languages? investigating large pretrained multilingual models for machine translation of Indian languages,” in *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, J. D. Pawar and S. Lalitha Devi, Eds. Goa University, Goa, India: NLP Association of India (NLP AI), Dec. 2023, pp. 663–669. [Online]. Available: <https://aclanthology.org/2023.icon-1.66/>
- [46] A. Galiano-Jiménez, F. Sánchez-Martínez, V. M. Sánchez-Cartagena, and J. A. Pérez-Ortiz, “Exploiting large pre-trained models for low-resource neural machine translation,” in *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, M. Nurminen, J. Brenner, M. Koponen, S. Latomaa, M. Mikhailov, F. Schierl, T. Ranasinghe, E. Vanmassenhove, S. A. Vidal, N. Aranberri, M. Nunziatini, C. P. Escartín,

- M. Forcada, M. Popovic, C. Scarton, and H. Moniz, Eds. Tampere, Finland: European Association for Machine Translation, Jun. 2023, pp. 59–68. [Online]. Available: <https://aclanthology.org/2023.eamt-1.7/>
- [47] I. J. Goodfellow, M. Mirza, D. Xiao, A. Courville, and Y. Bengio, “An empirical investigation of catastrophic forgetting in gradient-based neural networks,” 2015. [Online]. Available: <https://arxiv.org/abs/1312.6211>
- [48] Y. Liang, F. Meng, J. Xu, J. Wang, Y. Chen, and J. Zhou, “Unified model learning for various neural machine translation,” 2023. [Online]. Available: <https://arxiv.org/abs/2305.02777>
- [49] J. Wu, Y. Liu, and C. Zong, “F-malloc: Feed-forward memory allocation for continual learning in neural machine translation,” *arXiv preprint arXiv:2404.04846*, 2024.
- [50] D. Britz, Q. Le, and R. Pryzant, “Effective domain mixing for neural machine translation,” in *Proceedings of the Second Conference on Machine Translation*, O. Bojar, C. Buck, R. Chatterjee, C. Federmann, Y. Graham, B. Haddow, M. Huck, A. J. Yepes, P. Koehn, and J. Kreutzer, Eds. Copenhagen, Denmark: Association for Computational Linguistics, Sep. 2017, pp. 118–126. [Online]. Available: <https://aclanthology.org/W17-4712/>
- [51] Z. Liu, G. I. Winata, and P. Fung, “Continual mixed-language pre-training for extremely low-resource neural machine translation,” in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds. Online: Association for Computational Linguistics, Aug. 2021, pp. 2706–2718. [Online]. Available: <https://aclanthology.org/2021.findings-acl.239/>
- [52] W.-J. Ko, A. El-Kishky, A. Renduchintala, V. Chaudhary, N. Goyal, F. Guzmán, P. Fung, P. Koehn, and M. Diab, “Adapting high-resource NMT models to translate low-resource related languages without parallel data,” 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)*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds. Online: Association for Computational Linguistics, Aug. 2021, pp. 802–812. [Online]. Available: <https://aclanthology.org/2021.acl-long.66/>
- [53] A. Bapna and O. Firat, “Non-parametric adaptation for 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)*, J. Burstein, C. Doran, and

- T. Solorio, Eds. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 1921–1931. [Online]. Available: <https://aclanthology.org/N19-1191/>
- [54] R. Aharoni and Y. Goldberg, “Unsupervised domain clusters in pretrained language models,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds. Online: Association for Computational Linguistics, Jul. 2020, pp. 7747–7763. [Online]. Available: <https://aclanthology.org/2020.acl-main.692/>
- [55] C. Escolano, M. R. Costa-Jussà, and J. A. R. Fonollosa, “From bilingual to multilingual neural-based machine translation by incremental training,” *Journal of the Association for Information Science and Technology*, vol. 72, no. 2, pp. 190–203, 2021. [Online]. Available: <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.24395>
- [56] Y. Cao, H.-R. Wei, B. Chen, and X. Wan, “Continual learning for neural machine translation,” in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, and Y. Zhou, Eds. Online: Association for Computational Linguistics, Jun. 2021, pp. 3964–3974. [Online]. Available: <https://aclanthology.org/2021.naacl-main.310/>
- [57] B. Thompson, J. Gwinnup, H. Khayrallah, K. Duh, and P. Koehn, “Overcoming catastrophic forgetting during domain adaptation of 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)*, J. Burstein, C. Doran, and T. Solorio, Eds. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 2062–2068. [Online]. Available: <https://aclanthology.org/N19-1209/>
- [58] D. Saunders and S. DeNeefe, “Domain adapted machine translation: What does catastrophic forgetting forget and why?” in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 12 660–12 671. [Online]. Available: <https://aclanthology.org/2024.emnlp-main.704/>
- [59] E. Hasler, T. Domhan, J. Trenous, K. Tran, B. Byrne, and F. Hieber, “Improving the quality trade-off for neural machine translation multi-domain adaptation,” in *Proceedings of the 2021 Conference on Empirical Methods*

- in *Natural Language Processing*, M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, Eds. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 8470–8477. [Online]. Available: <https://aclanthology.org/2021.emnlp-main.666/>
- [60] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, “Overcoming catastrophic forgetting in neural networks,” *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, p. 3521–3526, Mar. 2017. [Online]. Available: <http://dx.doi.org/10.1073/pnas.1611835114>
- [61] J. Park, J. Song, and S. Yoon, “Building a neural machine translation system using only synthetic parallel data,” 2017. [Online]. Available: <https://arxiv.org/abs/1704.00253>
- [62] R. Sennrich, B. Haddow, and A. Birch, “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)*, K. Erk and N. A. Smith, Eds. Berlin, Germany: Association for Computational Linguistics, Aug. 2016, pp. 86–96. [Online]. Available: <https://aclanthology.org/P16-1009/>
- [63] S. Tars and M. Fishel, “Multi-domain neural machine translation,” *arXiv preprint arXiv:1805.02282*, 2018.
- [64] K. Song, Y. Zhang, H. Yu, W. Luo, K. Wang, and M. Zhang, “Code-switching for enhancing nmt with pre-specified translation,” 2019. [Online]. Available: <https://arxiv.org/abs/1904.09107>
- [65] Z. Yang, W. Chen, F. Wang, and B. Xu, “Generative adversarial training for neural machine translation,” *Neurocomputing*, vol. 321, pp. 146–155, 2018.
- [66] S. Tars and M. Fishel, “Multi-domain neural machine translation,” in *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, J. A. Pérez-Ortiz, F. Sánchez-Martínez, M. Esplà-Gomis, M. Popović, C. Rico, A. Martins, J. Van den Bogaert, and M. L. Forcada, Eds., Alicante, Spain, May 2018, pp. 279–288. [Online]. Available: <https://aclanthology.org/2018.eamt-main.26/>
- [67] H. Jiang, C. Liang, C. Wang, and T. Zhao, “Multi-domain neural machine translation with word-level adaptive layer-wise domain mixing,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds. Online: Association

- for Computational Linguistics, Jul. 2020, pp. 1823–1834. [Online]. Available: <https://aclanthology.org/2020.acl-main.165/>
- [68] J. Lee, H. Kim, H. Cho, E. Choi, and C. Park, “Specializing multi-domain NMT via penalizing low mutual information,” in *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, Y. Goldberg, Z. Kozareva, and Y. Zhang, Eds. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, Dec. 2022, pp. 10 015–10 026. [Online]. Available: <https://aclanthology.org/2022.emnlp-main.680/>
- [69] W. Wang, Y. Tian, J. Ngiam, Y. Yang, I. Caswell, and Z. Parekh, “Learning a multi-domain curriculum for neural machine translation,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds. Online: Association for Computational Linguistics, Jul. 2020, pp. 7711–7723. [Online]. Available: <https://aclanthology.org/2020.acl-main.689/>
- [70] A. Sharaf, H. Hassan, and H. Daumé III, “Meta-learning for few-shot NMT adaptation,” in *Proceedings of the Fourth Workshop on Neural Generation and Translation*, A. Birch, A. Finch, H. Hayashi, K. Heafield, M. Junczys-Dowmunt, I. Konstas, X. Li, G. Neubig, and Y. Oda, Eds. Online: Association for Computational Linguistics, Jul. 2020, pp. 43–53. [Online]. Available: <https://aclanthology.org/2020.ngt-1.5/>
- [71] E. Stergiadis, S. Kumar, F. Kovalev, and P. Levin, “Multi-domain adaptation in neural machine translation through multidimensional tagging,” in *Proceedings of Machine Translation Summit XVIII: Users and Providers Track*, J. Campbell, B. Huyck, S. Larocca, J. Marciano, K. Savenkov, and A. Yanishevsky, Eds. Virtual: Association for Machine Translation in the Americas, Aug. 2021, pp. 396–420. [Online]. Available: <https://aclanthology.org/2021.mtsummit-up.27/>
- [72] A. Cooper Stickland, A. Berard, and V. Nikoulina, “Multilingual domain adaptation for NMT: Decoupling language and domain information with adapters,” in *Proceedings of the Sixth Conference on Machine Translation*, L. Barrault, O. Bojar, F. Bougares, R. Chatterjee, M. R. Costa-jussa, C. Federmann, M. Fishel, A. Fraser, M. Freitag, Y. Graham, R. Grundkiewicz, P. Guzman, B. Haddow, M. Huck, A. J. Yepes, P. Koehn, T. Kocmi, A. Martins, M. Morishita, and C. Monz, Eds. Online: Association for Computational Linguistics, Nov. 2021, pp. 578–598. [Online]. Available: <https://aclanthology.org/2021.wmt-1.64/>
- [73] J. Guo, K. Han, Y. Wang, H. Wu, X. Chen, C. Xu, and C. Xu, “Distilling object

detectors via decoupled features,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2154–2164.

- [74] A. Moslemi, A. Briskina, Z. Dang, and J. Li, “A survey on knowledge distillation: Recent advancements,” *Machine Learning with Applications*, vol. 18, p. 100605, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666827024000811>
- [75] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, “Fitnets: Hints for thin deep nets,” 2015. [Online]. Available: <https://arxiv.org/abs/1412.6550>
- [76] J. Yim, D. Joo, J. Bae, and J. Kim, “A gift from knowledge distillation: Fast optimization, network minimization and transfer learning,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 7130–7138.
- [77] W. Park, D. Kim, Y. Lu, and M. Cho, “Relational knowledge distillation,” 2019. [Online]. Available: <https://arxiv.org/abs/1904.05068>
- [78] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Eds., vol. 27. Curran Associates, Inc., 2014. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2014/file/a14ac55a4f27472c5d894ec1c3c743d2-Paper.pdf
- [79] D. L. Silver, Q. Yang, and L. Li, “Lifelong machine learning systems: Beyond learning algorithms,” in *2013 AAAI spring symposium series*, 2013.
- [80] W. Jooste, R. Haque, and A. Way, “Knowledge distillation: A method for making neural machine translation more efficient,” *Information*, vol. 13, no. 2, p. 88, 2022.
- [81] Y. Li, C. Zhao, and C. Caragea, “Improving stance detection with multi-dataset learning and knowledge distillation,” in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, Eds. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 6332–6345. [Online]. Available: <https://aclanthology.org/2021.emnlp-main.511/>
- [82] A. Ganesan, F. Ferraro, and T. Oates, “Learning a reversible embedding mapping using bi-directional manifold alignment,” in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds. Online: Association for Computational Linguistics, Aug. 2021,

- pp. 3132–3139. [Online]. Available: <https://aclanthology.org/2021.findings-acl.276/>
- [83] S. Wu and M. Dredze, “Do explicit alignments robustly improve multilingual encoders?” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds. Online: Association for Computational Linguistics, Nov. 2020, pp. 4471–4482. [Online]. Available: <https://aclanthology.org/2020.emnlp-main.362/>
- [84] K. Huang, X. Guo, and M. Wang, “Towards efficient pre-trained language model via feature correlation distillation,” in *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36. Curran Associates, Inc., 2023, pp. 16 114–16 128. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/34260a400e39a802961470b3d3de99cc-Paper-Conference.pdf
- [85] P. O. Aboagye, Y. Zheng, M. Yeh, J. Wang, Z. Zhuang, H. Chen, L. Wang, W. Zhang, and J. Phillips, “Quantized Wasserstein Procrustes alignment of word embedding spaces,” in *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, K. Duh and F. Guzmán, Eds. Orlando, USA: Association for Machine Translation in the Americas, Sep. 2022, pp. 200–214. [Online]. Available: <https://aclanthology.org/2022.amta-research.15/>
- [86] S. Cao, N. Kitaev, and D. Klein, “Multilingual alignment of contextual word representations,” 2020. [Online]. Available: <https://arxiv.org/abs/2002.03518>
- [87] J. Zhang, C. Dong, X. Duan, Y. Zhang, and M. Zhang, “Third-party aligner for neural word alignments,” in *Findings of the Association for Computational Linguistics: EMNLP 2022*, Y. Goldberg, Z. Kozareva, and Y. Zhang, Eds. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, Dec. 2022, pp. 3134–3145. [Online]. Available: <https://aclanthology.org/2022.findings-emnlp.228/>
- [88] P. BehnamGhader, V. Adlakha, M. Mosbach, D. Bahdanau, N. Chapados, and S. Reddy, “Llm2vec: Large language models are secretly powerful text encoders,” 2024. [Online]. Available: <https://arxiv.org/abs/2404.05961>
- [89] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, and A. Rush, “Transformers: State-of-the-art natural language processing,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language*

- Processing: System Demonstrations*, Q. Liu and D. Schlangen, Eds. Online: Association for Computational Linguistics, Oct. 2020, pp. 38–45. [Online]. Available: <https://aclanthology.org/2020.emnlp-demos.6/>
- [90] J. Tiedemann, “Parallel data, tools and interfaces in OPUS,” in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, and S. Piperidis, Eds. Istanbul, Turkey: European Language Resources Association (ELRA), May 2012, pp. 2214–2218. [Online]. Available: <https://aclanthology.org/L12-1246/>
- [91] P. Koehn, “Europarl: A parallel corpus for statistical machine translation,” in *Proceedings of Machine Translation Summit X: Papers*, Phuket, Thailand, Sep. 13-15 2005, pp. 79–86. [Online]. Available: <https://aclanthology.org/2005.mtsummit-papers.11/>
- [92] P. Lison and J. Tiedemann, “OpenSubtitles2016: Extracting large parallel corpora from movie and TV subtitles,” in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, and S. Piperidis, Eds. Portorož, Slovenia: European Language Resources Association (ELRA), May 2016, pp. 923–929. [Online]. Available: <https://aclanthology.org/L16-1147/>
- [93] N. Reimers and I. Gurevych, “Making monolingual sentence embeddings multilingual using knowledge distillation,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds. Online: Association for Computational Linguistics, Nov. 2020, pp. 4512–4525. [Online]. Available: <https://aclanthology.org/2020.emnlp-main.365/>
- [94] NLLB Team, M. R. Costa-jussà, J. Cross, O. Çelebi, M. Elbayad, K. Heafield, K. Heffernan, E. Kalbassi, J. Lam, D. Licht, J. Maillard, A. Sun, S. Wang, G. Wenzek, A. Youngblood, B. Akula, L. Barrault, G. M. Gonzalez, P. Hansanti, J. Hoffman, S. Jarrett, K. R. Sadagopan, D. Rowe, S. Spruit, C. Tran, P. Andrews, N. F. Ayan, S. Bhosale, S. Edunov, A. Fan, C. Gao, V. Goswami, F. Guzmán, P. Koehn, A. Mourachko, C. Ropers, S. Saleem, H. Schwenk, and J. Wang, “Scaling neural machine translation to 200 languages,” *Nature*, vol. 630, no. 8018, pp. 841–846, 2024. [Online]. Available: <https://doi.org/10.1038/s41586-024-07335-x>
- [95] C. Raffel, N. M. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, “Exploring the limits of transfer learning with a unified

text-to-text transformer,” *J. Mach. Learn. Res.*, vol. 21, pp. 140:1–140:67, 2019.
[Online]. Available: <https://api.semanticscholar.org/CorpusID:204838007>