

References

- [1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need,” in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *arXiv preprint arXiv:1810.04805*, 2018.
- [3] F. A. Sheikha and D. Inkpen, “Generation of Formal and Informal Sentences,” in *Proceedings of the 13th European Workshop on Natural Language Generation*, 2011, pp. 187–193.
- [4] S. Rao and J. R. Tetreault, “Dear Sir or Madam, May I Introduce the GYAFC Dataset: Corpus, Benchmarks and Metrics for Formality Style Transfer,” *CoRR*, vol. abs/1803.06535, 2018. [Online]. Available: <http://arxiv.org/abs/1803.06535>
- [5] X. Niu, S. Rao, and M. Carpuat, “Multi-Task Neural Models for Translating Between Styles Within and Across Languages,” *arXiv preprint arXiv:1806.04357*, 2018.
- [6] R. Xu, T. Ge, and F. Wei, “Formality Style Transfer with Hybrid Textual Annotations,” *arXiv preprint arXiv:1903.06353*, 2019.
- [7] Y. Wang, Y. Wu, L. Mou, Z. Li, and W. Chao, “Harnessing Pre-Trained Neural Networks with Rules for Formality Style Transfer,” 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, pp. 3564–3569.
- [8] M. Malmsten, L. Börjeson, and C. Haffenden, “Playing with Words at the National Library of Sweden—Making a Swedish BERT,” *arXiv preprint arXiv:2007.01658*, 2020.

- [9] “Hejare AB,” URL: <https://www.hejare.se/>, Last accessed 28 January 2021.
- [10] “Momang,” URL: <https://momang.io/>, Last accessed 28 January 2021.
- [11] F. Heylighen and J.-M. Dewaele, “Formality of Language: definition, measurement and behavioral determinants,” *Interner Bericht, Center “Leo Apostel”, Vrije Universiteit Brüssel*, vol. 4, 1999.
- [12] E. Hovy, “Generating Natural Language under Pragmatic Constraints,” *Journal of Pragmatics*, vol. 11, no. 6, pp. 689–719, 1987.
- [13] E. Pavlick and J. Tetreault, “An Empirical Analysis of Formality in Online Communication,” *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 61–74, 2016.
- [14] W. Xu, A. Ritter, W. B. Dolan, R. Grishman, and C. Cherry, “Paraphrasing for Style,” in *Proceedings of COLING 2012*, 2012, pp. 2899–2914.
- [15] Z. Hu, Z. Yang, X. Liang, R. Salakhutdinov, and E. P. Xing, “Toward Controlled Generation of Text,” in *International Conference on Machine Learning*. PMLR, 2017, pp. 1587–1596.
- [16] “Yahoo Datasets,” URL: <https://webscope.sandbox.yahoo.com/catalog.php?datatype=l>, Last accessed 21 February 2021.
- [17] H. Jhamtani, V. Gangal, E. Hovy, and E. Nyberg, “Shakespearizing Modern Language Using Copy-Enriched Sequence-to-Sequence Models,” *arXiv preprint arXiv:1707.01161*, 2017.
- [18] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, “Language Models are Unsupervised Multitask Learners,” *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.
- [19] “Yahoo! Answers,” URL: <https://answers.yahoo.com/answer>, Last accessed 21 February 2021.
- [20] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “BLEU: a Method for Automatic Evaluation of Machine Translation,” in *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002, pp. 311–318.

- [21] D. Chen and W. B. Dolan, “Collecting Highly Parallel Data for Paraphrase Evaluation,” in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011, pp. 190–200.
- [22] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” *arXiv preprint arXiv:1301.3781*, 2013.
- [23] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey *et al.*, “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation,” *arXiv preprint arXiv:1609.08144*, 2016.
- [24] T. Kudo and J. Richardson, “SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing,” *arXiv preprint arXiv:1808.06226*, 2018.
- [25] F. Rosenblatt, “Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms,” Cornell Aeronautical Lab Inc Buffalo NY, Tech. Rep., 1961.
- [26] J. J. Hopfield, “Neural Networks and Physical Systems with Emergent Collective Computational Abilities,” *Proceedings of the national academy of sciences*, vol. 79, no. 8, pp. 2554–2558, 1982.
- [27] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [28] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to Sequence Learning with Neural Networks,” *arXiv preprint arXiv:1409.3215*, 2014.
- [29] 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,” *arXiv preprint arXiv:1406.1078*, 2014.
- [30] D. Bahdanau, K. Cho, and Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” *arXiv preprint arXiv:1409.0473*, 2014.
- [31] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, “Deep contextualized word representations,” *arXiv preprint arXiv:1802.05365*, 2018.

- [32] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, “Improving Language Understanding with Unsupervised Learning,” *Technical report, OpenAI*, 2018.
- [33] Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler, “Aligning Books and Movies: Towards Story-like Visual Explanations by Watching Movies and Reading Books,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 19–27.
- [34] T. Pires, E. Schlinger, and D. Garrette, “How multilingual is Multilingual BERT?” *arXiv preprint arXiv:1906.01502*, 2019.
- [35] S. Wu and M. Dredze, “Beto, Bentz, Becas: The Surprising Cross-Lingual Effectiveness of BERT,” *arXiv preprint arXiv:1904.09077*, 2019.
- [36] S. Rothe, S. Narayan, and A. Severyn, “Leveraging Pre-trained Checkpoints for Sequence Generation Tasks,” *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 264–280, 2020.
- [37] “Google Scholar,” URL: <https://scholar.google.com/>, Last accessed 28 April 2021.
- [38] “Google Cloud,” URL: <https://cloud.google.com>, Last accessed 28 January 2021.
- [39] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, “ALBERT: A Lite BERT for Self-supervised Learning of Language Representations,” *arXiv preprint arXiv:1909.11942*, 2019.
- [40] “Kb/bert-base-swedish-cased, Huggingface Model,” URL: <https://huggingface.co/KB/bert-base-swedish-cased>, Last accessed 16 April 2021.
- [41] “Python,” URL: <https://www.python.org/>, Last accessed 28 January 2021.
- [42] “Transformers, Huggingface Library,” URL: <https://huggingface.co/transformers/>, Last accessed 16 April 2021.
- [43] “EncoderDecoderModel, Huggingface Library,” URL: https://huggingface.co/transformers/model_doc/encoderdecoder.html#encoderdecodermodel, Last accessed 16 April 2021.

- [44] “bert-base-cased, Huggingface Model,” URL: <https://huggingface.co/bert-base-cased>, Last accessed 16 April 2021.
- [45] “BertTokenizer, Huggingface Library,” URL: https://huggingface.co/transformers/model_doc/bert.html#berttokenizer, Last accessed 16 April 2021.
- [46] “Seq2SeqTrainer, Huggingface Library,” URL: https://huggingface.co/transformers/main_classes/trainer.html#seq2seqtrainer, Last accessed 16 April 2021.
- [47] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [48] “Tensorflow,” URL: <https://www.tensorflow.org/>, Last accessed 28 January 2021.
- [49] “PyTorch,” URL: <https://pytorch.org/>, Last accessed 28 January 2021.
- [50] “Apex (A PyTorch Extension), NVIDIA,” URL: <https://nvidia.github.io/apex/>, Last accessed 16 April 2021.
- [51] “Debian,” URL: <https://www.debian.org/>, Last accessed 16 April 2021.
- [52] “Deep Learning VM, Google Cloud,” URL: <https://cloud.google.com/deep-learning-vm>, Last accessed 16 April 2021.
- [53] “NVIDIA Tesla T4 GPU,” URL: <https://www.nvidia.com/en-us/data-center/tesla-t4/>, Last accessed 16 April 2021.
- [54] 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,” *arXiv preprint arXiv:2010.11934*, 2020.
- [55] “GYAFC Github,” URL: <https://github.com/raosudha89/GYAFC-corpus>, Last accessed 28 April 2021.

