

## DAFTAR PUSTAKA

- Abas, H., Shaidy, & Jaya, L., 1985, *Kamus Wolio-Indonesia*, Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan, Jakarta.
- Abas, Husen, Manyambeang, A. Kadir, Nandar, Ibnu, Shaidy, 1983, *Struktur Bahasa Wolio*, Pusat Pembinaan dan Pengembangan Bahasa Departemen Pendidikan dan Kebudayaan, Jakarta.
- Aji, A. F., Winata, G. I., Koto, F., Cahyawijaya, S., Romadhony, A., Mahendra, R., Kurniawan, K., Moeljadi, D., Prasojo, R. E., Baldwin, T., Lau, J. H., & Ruder, S., 2022, One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia, *Proceedings Of the 60th Annual Meeting of the Association for Computational Linguistics*, <https://doi.org/10.18653/v1/2022.acl-long.500>, 7226-7249.
- Alammar, J., 2018, The Illustrated Transformer : Visualizing machine learning one concept at a time, <https://jalammar.github.io/illustrated-transformer/>, 27 Juni 2018, diakses 03 Desember 2023.
- Ameur, M. S. H., Meziane, F., & Guessoum, A., 2020, Arabic Machine Translation: A survey of the latest trends and challenges, *Computer Science Review*, <https://doi.org/10.1016/j.cosrev.2020.100305>, 38, 100305.
- Araabi, A., & Monz, C., 2020, Optimizing Transformer for Low-Resource Neural Machine Translation, *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, <https://doi.org/10.18653/v1/2020.coling-main.304>, 3429-3435.
- Bahdanau, D., Cho, K. H., & Bengio, Y., 2015, Neural machine translation by jointly learning to align and translate, *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, <https://arxiv.org/abs/1409.0473>, 1-15.
- Bawden, R., Bogoychev, N., Germann, U., Grundkiewicz, R., Kirefu, F., Barone, A. V. M., & Birch, A., 2019, The University of Edinburgh's Submissions to the WMT19 News Translation Task, *WMT 2019 - 4th Conference on Machine Translation, Proceedings of the Conference*, <https://doi.org/10.18653/v1/w19-5347>, 2(Day 1), 412-423.
- Baziotis, C., Haddow, B., & Birch, A., 2020, Language model prior for low-resource neural machine translation, *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, <https://doi.org/10.18653/v1/2020.emnlp-main.615>, 7622-7634.
- Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L., 1993, The Mathematics of Statistical Machine Translation: Parameter Estimation, *Computational Linguistics*, <https://aclanthology.org/J93-2003.pdf>,

10598(August), 263–311.

- Burchell, L., Birch, A. and Heafield, K., 2022, Exploring Diversity in Back Translation for Low-Resource Machine Translation, *DeepLo 2022 - 3rd Workshop on Deep Learning Approaches for Low-Resource NLP*, doi: 10.18653/v1/2022.deeplo-1.8, Proceedings of the DeepLo Workshop, pp. 67–79.
- Caswell, I., Chelba, C., & Grangier, D., 2019, Tagged back-translation, *WMT 2019 - 4th Conference on Machine Translation, Proceedings of the Conference*, <https://doi.org/10.18653/v1/w19-5206>, 53–63.
- Chen, P. J., Lee, A., Wang, C., Goyal, N., Fan, A., Williamson, M., & Gu, J., 2020, Facebook AI’s WMT20 News Translation Task Submission, *WMT 2020 - 5th Conference on Machine Translation*, <https://aclanthology.org/2020.wmt-1.8>, Proceedings, 205–215.
- Dandapat, S., & Federmann, C., 2018, Iterative Data Augmentation for Neural Machine Translation: a Low Resource Case Study for English–Telugu, *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, <https://aclanthology.org/2018.eamt-main.29>, May, 307–312.
- Data Pokok Kebahasaan dan Kesastraan, Kemdikbud, 2019, Persebaran Bahasa Daerah Berdasarkan Provinsi, [dapobas.kemdikbud.go.id/home?show=isidata&id=287](http://dapobas.kemdikbud.go.id/home?show=isidata&id=287), 17 Agustus, diakses 17 November 2024.
- Ding, S., Renduchintala, A., & Duh, K., 2019, A Call for Prudent Choice of Subword Merge Operations in Neural Machine Translation, *Proceedings of Machine Translation Summit XVII: Research Track*, <https://aclanthology.org/W19-6620>, 204–213.
- Djabaru, L. O., 2017, *Kamus Bahasa Wolio (Buton)*, Ghina Pustaka, Makassar.
- Dou, Z. Y., Anastasopoulos, A., & Neubig, G., 2020, Dynamic data selection and weighting for iterative back-translation, *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, <https://doi.org/10.18653/v1/2020.emnlp-main.475>, 5894–5904.
- Eberhard, D. M., F. Simons, G., & D. Fennig, C., 2024, *Ethnologue: Languages of the World (Wolio)*, Twenty-seventh edition. Dallas, Texas: SIL International, Online version: <http://www.ethnologue.com>
- Edunov, S., Ott, M., Auli, M., & Grangier, D., 2018, Understanding back-translation at scale, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, <https://doi.org/10.18653/v1/d18-1045>, EMNLP 2018, 489–500.
- Gamal, D., Alfonse, M., Jimenez-Zafra, S. M., & Aref, M., 2022, Survey of Arabic Machine Translation, Methodologies, Progress, and Challenges, *MIUCC 2022 - 2nd International Mobile, Intelligent, and Ubiquitous*

*Computing Conference*, 10.1109/MIUCC55081.2022.9781776, 378-383.

- Garcia, X., Siddhant, A., Firat, O., & Parikh, A. P., 2021, Harnessing Multilinguality in Unsupervised Machine Translation for Rare Languages, *NAACL-HLT 2021 - 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, <https://doi.org/10.18653/v1/2021.naacl-main.89>, Figure 1, 1126–1137.
- Goel, P., Prabhu, S., Debnath, A., Modi, P., & Shrivastava, M., 2020, Hindi TimeBank: An ISO-TimeML Annotated Reference Corpus, *16th Joint ACL - ISO Workshop on Interoperable Semantic Annotation PROCEEDINGS*, <https://aclanthology.org/2020.isa-1.2>, May, 13–21.
- Goyal, N., Gao, C., Chaudhary, V., Chen, P. J., Wenzek, G., Ju, D., Krishnan, S., Ranzato, M., Guzmán, F., & Fan, A., 2022, The FLORES-101 Evaluation Benchmark for Low-Resource and Multilingual Machine Translation, *Transactions of the Association for Computational Linguistics*, [https://doi.org/10.1162/tacl\\_a\\_00474](https://doi.org/10.1162/tacl_a_00474), 10, 522–538.
- Gulcehre, C., Firat, O., Xu, K., Cho, K., Barrault, L., Lin, H.-C., Bougares, F., Schwenk, H., & Bengio, Y., 2015, On Using Monolingual Corpora in Neural Machine Translation, *CoRR*, <http://arxiv.org/abs/1503.03535>, abs/1503.03535.
- Guntara, T. W., Aji, A. F., & Prasajo, R. E. , 2020, Benchmarking Multidomain English-Indonesian Machine Translation, *Proceedings of the 13th Workshop on Building and Using Comparable Corpora*, <https://aclanthology.org/2020.bucc-1.6>, May, 35–43.
- Guzmán, F., Chen, P. J., Ott, M., Pino, J., Lample, G., Koehn, P., Chaudhary, V., & Ranzato, M., 2019, The Flores evaluation datasets for low-resource machine translation: Nepali-English and Sinhala-English, *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*, <https://doi.org/10.18653/v1/d19-1632>, Proceedings of the Conference, 6098–6111.
- Haddow, B., Bawden, R., Barone, A. V. M., Helcl, J., & Birch, A., 2022, Survey of Low-Resource Machine Translation, *Computational Linguistics*, [https://doi.org/10.1162/coli\\_a\\_00446](https://doi.org/10.1162/coli_a_00446), 48(3), 673–732.
- Hlaing, Z. Z., Thu, Y. K., Supnithi, T., & Netisopakul, P., 2022, Improving neural machine translation with POS-tag features for low-resource language pairs, *Heliyon*, <https://doi.org/10.1016/j.heliyon.2022.e10375>, 8(8), e10375.
- Hoang, V. C. D., Koehn, P., Haffari, G., & Cohn, T. , 2018, Iterative Back-Translation for Neural Machine Translation, *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*,

- <https://doi.org/10.18653/v1/w18-2703>, 2016, 18–24.
- Jain, A., 2024, TF-IDF Vectorization with Cosine Similarity, <https://medium.com/@anurag-jain/tf-idf-vectorization-with-cosine-similarity-eca3386d4423>, 4 Februari 2024, diakses 2 Desember 2024.
- Jurafsky, Daniel and Martin, James H, 2023, *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition Third Edition draft*, Stanford University, California
- Klein, G., Hernandez, F., Nguyen, V., & Senellart, J, 2020, The OpenNMT Neural Machine Translation Toolkit: 2020 Edition, *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*, <https://opennmt.net>, 1, 102–109.
- Koehn, P., Och, F. J., & Marcu, D., 2017, Six Challenges for Neural Machine Translation, *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, <https://doi.org/10.18653/v1/w17-3204>, 28–39.
- Koehn, P., Och, F. J., & Marcu, D., 2003, Statistical Phrase-Based Translation, *Proceedings of the Joint Conference on Human Language Technologies and the Annual Meeting of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL)*, <https://doi.org/10.1007/BF00293403>, 88(3), 262–266.
- Koehn, Philipp, 2020, *Neural machine translation*, Cambridge University Press, Cambridge.
- Kostadinov, Simeon, 2017, Understanding GRU Networks, <https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be>, 16 Desember, diakses 16 Januari 2025.
- Koto, Fajri & Koto, Ikhwan, 2020, Towards Computational Linguistics in Minangkabau Language: Studies on Sentiment Analysis and Machine Translation, *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, <https://aclanthology.org/2020.paclic-1.17>, 138--148.
- Kudo, T., 2018, Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, <https://doi.org/10.18653/v1/P18-1007>, 66-75.
- Lai, W., Libovický, J., & Fraser, A., 2021, The LMU Munich System for the WMT 2021 Large-Scale Multilingual Machine Translation Shared Task, *WMT 2021 - 6th Conference on Machine Translation, Proceedings*, <https://aclanthology.org/2021.wmt-1.49.pdf>, 412–417.
- Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M., & Zettlemoyer, L., 2020, Multilingual denoising pre-training for neural

- machine translation, *Transactions of the Association for Computational Linguistics*, [https://doi.org/10.1162/tacl\\_a\\_00343](https://doi.org/10.1162/tacl_a_00343), 8, 726–742.
- Luong, M.-T., Pham, H., & D. Manning, C., 2016, Effective Approaches to Attention-based Neural Machine Translation, *Proceedings Of the 2015 Conference on Empirical Methods in Natural Language Processing*, [https://doi.org/10.1007/978-3-319-28308-1\\_29](https://doi.org/10.1007/978-3-319-28308-1_29), September, 471–482.
- Markowitz, Dale, 2021, Transformers, Explained: Understand the Model Behind GPT-3, BERT, and T5, <https://daleonai.com/transformers-explained>, 6 Mei, diakses 27 November 2024.
- Moeliono, A. M., Lapoliwa, H., Alwi, H., Tjatur Wisnu Sasangka, S. S., & Sugiyono, 2017, *Tata Bahasa Baku Bahasa Indonesia Edisi Keempat*, Badan Pengembangan dan Pembinaan Bahasa Kementerian Pendidikan dan Kebudayaan, Jakarta
- Mosolova, A., & Smaili, K., 2022, The Only Chance to Understand: Machine Translation of the Severely Endangered Low-resource Languages of Eurasia, *Proceedings of the Fifth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2022)*, <https://aclanthology.org/2022.loresmt-1.4>, 23–34.
- Moukafih, Y., Sbihi, N., Ghogho, M., Smaïli, K., Moukafih, Y., Sbihi, N., Ghogho, M., Smaïli, K., & Machine, I., 2021, Improving Machine Translation of Arabic Dialects through Multi-Task Learning, *20th International Conference Italian Association for Artificial Intelligence*, [http://dx.doi.org/10.1007/978-3-031-08421-8\\_40](http://dx.doi.org/10.1007/978-3-031-08421-8_40), AIXIA.
- Munteanu, Dragos Stefan & Marcu, Daniel, 2005, Improving Machine Translation Performance by Exploiting Non-Parallel Corpora, *Computational Linguistics*, <https://doi.org/10.1162/089120105775299168>, 477-504.
- Neubig, G., 2017, Neural Machine Translation and Sequence-to-sequence Models: A Tutorial, *pp*, <http://arxiv.org/abs/1703.01619>, 1–65.
- Olah, Christopher, 2015, Understanding LSTM Networks, <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>, 27 Agustus, diakses 16 Januari 2025.
- Papineni, K., Roukos, S., Ward, T., Zhu, W., , BLEU: a method for automatic evaluation of machine translation, *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02)*, <https://doi.org/10.3115/1073083.1073135>, 311–318.
- Patel, Ankur A., and Arasanipalai, Ajay U., 2021, Applied Natural Language Processing in the Enterprise: Chapter 4. Tokenization, <https://www.oreilly.com/library/view/applied-natural-language/9781492062561/ch04.html>, 15 Juni, diakses 06 Mei 2024.

- Popović, M., 2015, chrF: character n-gram F-score for automatic MT evaluation, *Proceedings of the Tenth Workshop on Statistical Machine Translation*, <https://doi.org/10.18653/v1/W15-3049>, 392–395.
- Popović, M., 2015, chrF: character n-gram F-score for automatic MT evaluation, *Proceedings of the Tenth Workshop on Statistical Machine Translation*, <https://doi.org/10.18653/v1/W15-3049>, 392–395.
- Prijono, B., 2018, Pengenalan Recurrent Neural Network (RNN), <https://indoml.com/2018/04/04/pengenalan-rnn-bag-1/>, 4 April 2018, diakses 1 Desember 2024.
- Provilkov, I., Emelianenko, D., & Voita, E., 2020, BPE-dropout: Simple and effective subword regularization, *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, <https://doi.org/10.18653/v1/2020.acl-main.170>, 1, 1882–1892.
- Provilkov, I., Emelianenko, D., & Voita, E., 2020, BPE-dropout: Simple and effective subword regularization, *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, <https://doi.org/10.18653/v1/2020.acl-main.170>, 1, 1882–1892.
- Ranathunga, S., Lee, E.-S. A., Prifti Skenduli, M., Shekhar, R., Alam, M., & Kaur, R., , Neural Machine Translation for Low-Resource Languages: A Survey, *ACM Computing Surveys*, <https://doi.org/10.1145/3567592>, 55(11).
- Saleva, J., & Lignos, C., 2023, What changes when you randomly choose BPE merge operations? Not much, *The Fourth Workshop on Insights from Negative Results in NLP*, <https://doi.org/10.18653/v1/2023.insights-1.7>, 59–66.
- Schmidt, Robin, 2019, Recurrent Neural Networks (RNNs): A gentle Introduction and Overview, *CoRR*, <http://arxiv.org/abs/1912.05911>, abs/1912.05911.
- Sennrich, R., & Zhang, B, 2019, Revisiting low-resource neural machine translation: A case study, *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics*, <https://doi.org/10.18653/v1/p19-1021>, Proceedings of the Conference, 211–221.
- Sennrich, R., Haddow, B., & Birch, A., 2016a, Neural Machine Translation of Rare Words with Subword Units, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, <https://doi.org/10.18653/v1/P16-1162>, Volume 1.
- Sennrich, R., Haddow, B., & Birch, A., 2016b, Improving neural machine translation models with monolingual data, *54th Annual Meeting of the Association for Computational Linguistics*, <https://doi.org/10.18653/v1/p16-1009>, ACL 2016 - Long Papers, 1, 86–96.
- Shen, J., Chen, P. J., Le, M., He, J., Gu, J., Ott, M., Auli, M., & Ranzato, M., 2021, The source-target domain mismatch problem in machine translation,

- EACL 2021 - 16th Conference of the European Chapter of the Association for Computational Linguistics*, <https://doi.org/10.18653/v1/2021.eacl-main.130>, Proceedings of the Conference, 1519–1533.
- Stahlberg, F., Cross, J., & Stoyanov, V., 2018, Simple Fusion: Return of the Language Model, *WMT 2018 - 3rd Conference on Machine Translation, Proceedings of the Conference*, <https://doi.org/10.18653/v1/w18-6321>, 1, 204–211.
- Sutskever, I., Vinyals, O., & Le, Q. V., 2023, Translation Performance from the User’s Perspective of Large Language Models and Neural Machine Translation Systems, *Information (Switzerland)*, <https://doi.org/10.3390/info14100574>, 14(10).
- Sutskever, I., Vinyals, O., & Le, Q. V., 2014, Sequence to sequence learning with neural networks, *Advances in Neural Information Processing Systems*, <https://doi.org/10.48550/arXiv.1409.3215>, 4 (January), 3104–3112.
- Suyanto, Ramadhani, K. N., & Mandala, S., 2019, *Deep learning : modernisasi machine learning untuk big data*, Informatika, Bandung
- Tang, Y., Tran, C., Li, X., Chen, P. J., Goyal, N., Chaudhary, V., Gu, J., & Fan, A., 2021, Multilingual Translation from Denoising Pre-Training, *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, <https://doi.org/10.18653/v1/2021.findings-acl.304>, 3450–3466.
- Uszkoreit, Jacob, 2017, Transformer: A Novel Neural Network Architecture for Language Understanding, <https://research.google/blog/transformer-a-novel-neural-network-architecture-for-language-understanding/>, 31 Agustus, diakses 06 Mei 2023.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I., 2017, Attention is all you need, *Advances in Neural Information Processing Systems*, [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf), 2017-Decem(Nips), 5999–6009.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I., 2017, Attention is all you need, *Advances in Neural Information Processing Systems*, [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf), 2017-Decem(Nips), 5999–6009.
- Winata, G. I., Aji, A. F., Cahyawijaya, S., Mahendra, R., Koto, F., Romadhony, A., Kurniawan, K., Moeljadi, D., Prasajo, R. E., Fung, P., Baldwin, T., Lau, J. H., Sennrich, R., & Ruder, S., 2023, NusaX: Multilingual Parallel Sentiment Dataset for 10 Indonesian Local Languages, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, [10.18653/v1/2023.eacl-main.57](https://doi.org/10.18653/v1/2023.eacl-main.57), 815–834.

- Zhang, B., Xiong, D., Su, J., Duan, H., & Zhang, M., 2016, Variational neural machine translation, *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing*, <https://doi.org/10.18653/v1/d16-1050>, Proceedings, 521–530.
- Zouhar, V., Meister, C., Gastaldi, J. L., Du, L., Vieira, T., Sachan, M., & Cotterell, R., 2023, A Formal Perspective on Byte-Pair Encoding, *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, <https://doi.org/10.18653/v1/2023.findings-acl.38>, 598–614.
- Sanchez-Martinez, F., Sánchez-Cartagena, V. M., Pérez-Ortiz, J. A., Forcada, M. L., Esplà-Gomis, M., Secker, A., Coleman, S., & Wall, J., 2020, An English-Swahili parallel corpus and its use for neural machine translation in the news domain, *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, <https://aclanthology.org/2020.eamt-1.32>, EAMT 2020, 299–308.