UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR PUSTAKA

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., and Ruder, S. (2022). One country, 700+ languages: NLP challenges for underrepresented languages and dialects in Indonesia. In Muresan, S., Nakov, P., and Villavicencio, A., editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7226–7249, Dublin, Ireland. Association for Computational Linguistics.

Atkinson, R. C. and Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, 1(2):126–133.

Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., Do, Q. V., Xu, Y., and Fung, P. (2023). A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. In Park, J. C., Arase, Y., Hu, B., Lu, W., Wijaya, D., Purwarianti, A., and Krisnadhi, A. A., editors, *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 675–718, Nusa Dua, Bali. Association for Computational Linguistics.

Baziotis, C., Haddow, B., and Birch, A. (2020). Language model prior for low-resource neural machine translation. In Webber, B., Cohn, T., He, Y., and Liu, Y., editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7622–7634, Online. Association for Computational Linguistics.

Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss,

UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Cahyawijaya, S., Lovenia, H., Koto, F., Adhista, D., Dave, E., Oktavianti, S., Akbar, S., Lee, J., Shadieq, N., Cenggoro, T. W., Linuwih, H., Wilie, B., Muridan, G., Winata, G., Moeljadi, D., Aji, A. F., Purwarianti, A., and Fung, P. (2023a). NusaWrites: Constructing high-quality corpora for underrepresented and extremely low-resource languages. In Park, J. C., Arase, Y., Hu, B., Lu, W., Wijaya, D., Purwarianti, A., and Krisnadhi, A. A., editors, *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 921–945, Nusa Dua, Bali. Association for Computational Linguistics.

Cahyawijaya, S., Lovenia, H., Koto, F., Putri, R., Cenggoro, W., Lee, J., Akbar, S., Dave, E., Nuurshadieq, N., Mahendra, M., Putri, R., Wilie, B., Winata, G., Aji, A., Purwarianti, A., and Fung, P. (2024). Cendol: Open instruction-tuned generative large language models for Indonesian languages. In Ku, L.-W., Martins, A., and Srikumar, V., editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14899–14914, Bangkok, Thailand. Association for Computational Linguistics.

Cahyawijaya, S., Lovenia, H., Yu, T., Chung, W., and Fung, P. (2023b). InstructAlign: High-and-low resource language alignment via continual crosslingual instruction tuning. In Wijaya, D., Aji, A. F., Vania, C., Winata, G. I., and Purwarianti, A., editors, *Proceedings of the First Workshop in South East Asian Language Processing*, pages 55–78, Nusa Dua, Bali, Indonesia. Association for Computational Linguistics.

Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., and Jatowt, A. (2020). Yake! keyword extraction from single documents using multiple local features. *Information Sciences*, 509:257–289.

UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Moschitti, A., Pang, B., and Daelemans, W., editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar. Association for Computational Linguistics.

Chronopoulou, A., Stojanovski, D., and Fraser, A. (2021). Improving the lexical ability of pretrained language models for unsupervised neural machine translation. In Toutanova, K., Rumshisky, A., Zettlemoyer, L., Hakkani-Tur, D., Beltagy, I., Bethard, S., Cotterell, R., Chakraborty, T., and Zhou, Y., editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 173–180, Online. Association for Computational Linguistics.

Dabre, R., Chu, C., and Kunchukuttan, A. (2020). A survey of multilingual neural machine translation. *ACM Comput. Surv.*, 53(5).

Dang, J., Singh, S., D'souza, D., Ahmadian, A., Salamanca, A., Smith, M., Peppin, A., Hong, S., Govindassamy, M., Zhao, T., Kublik, S., Amer, M., Aryabumi, V., Campos, J. A., Tan, Y.-C., Kocmi, T., Strub, F., Grinsztajn, N., Flet-Berliac, Y., Locatelli, A., Lin, H., Talupuru, D., Venkitesh, B., Cairuz, D., Yang, B., Chung, T., Ko, W.-Y., Shi, S. S., Shukayev, A., Bae, S., Piktus, A., Castagné, R., Cruz-Salinas, F., Kim, E., Crawhall-Stein, L., Morisot, A., Roy, S., Blunsom, P., Zhang, I., Gomez, A., Frosst, N., Fadaee, M., Ermis, B., Üstün, A., and Hooker, S. (2024). Aya expanse: Combining research breakthroughs for a new multilingual frontier.

Dou, L., Liu, Q., Zeng, G., Guo, J., Zhou, J., Mao, X., Jin, Z., Lu, W., and Lin, M. (2024). Sailor: Open language models for south-East Asia. In Hernandez Farias, D. I., Hope, T., and Li, M., editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 424–435, Miami, Florida, USA. Association for Computational Linguistics.

Duan, X., Ji, B., Jia, H., Tan, M., Zhang, M., Chen, B., Luo, W., and Zhang, Y. (2020). Bilingual dictionary based neural machine translation without using parallel sentences. In Jurafsky, D., Chai, J., Schluter, N., and Tetreault, J., editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1570–1579, Online. Association for Computational Linguistics.

UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

Dyer, C., Chahuneau, V., and Smith, N. A. (2013). A Simple, Fast, and Effective Reparameterization of IBM Model 2. In Vanderwende, L., Daumé III, H., and Kirchhoff, K., editors, *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648, Atlanta, Georgia. Association for Computational Linguistics.

Feng, T., Wang, Z., and Sun, J. (2023). Citing: Large language models create curriculum for instruction tuning.

Gehring, J., Auli, M., Grangier, D., and Dauphin, Y. (2017). A convolutional encoder model for neural machine translation. In Barzilay, R. and Kan, M.-Y., editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 123–135, Vancouver, Canada. Association for Computational Linguistics.

Ghazvininejad, M., Gonen, H., and Zettlemoyer, L. (2023). Dictionary-based phrase-level prompting of large language models for machine translation.

Graves, M. F. (1986). Vocabulary learning and instruction. *Review of Research in Education*, 13:49.

Grootendorst, M. (2020). Keybert: Minimal keyword extraction with bert.

Guo, P., Ren, Y., Hu, Y., Li, Y., Zhang, J., Zhang, X., and Huang, H. (2024). Teaching large language models to translate on low-resource languages with textbook prompting. In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 15685–15697, Torino, Italia. ELRA and ICCL.

Haddow, B., Bawden, R., Miceli Barone, A. V., Helcl, J., and Birch, A. (2022). Survey of low-resource machine translation. *Computational Linguistics*, 48(3):673–732.

He, Z., Liang, T., Jiao, W., Zhang, Z., Yang, Y., Wang, R., Tu, Z., Shi, S., and Wang, X. (2024). Exploring human-like translation strategy with large language models. *Transactions of the Association for Computational Linguistics*, 12:229–246.

UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

Hedderich, M. A., Lange, L., Adel, H., Strötgen, J., and Klakow, D. (2021). A survey on recent approaches for natural language processing in low-resource scenarios. In Toutanova, K., Rumshisky, A., Zettlemoyer, L., Hakkani-Tur, D., Beltagy, I., Bethard, S., Cotterell, R., Chakraborty, T., and Zhou, Y., editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2545–2568, Online. Association for Computational Linguistics.

Heffernan, K., Çelebi, O., and Schwenk, H. (2022). Bitext mining using distilled sentence representations for low-resource languages. In Goldberg, Y., Kozareva, Z., and Zhang, Y., editors, *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 2101–2112, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Huang, H., Tang, T., Zhang, D., Zhao, X., Song, T., Xia, Y., and Wei, F. (2023). Not all languages are created equal in LLMs: Improving multilingual capability by cross-lingual-thought prompting. In Bouamor, H., Pino, J., and Bali, K., editors, *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 12365–12394, Singapore. Association for Computational Linguistics.

Hughes, C. A., Morris, J. R., Therrien, W. J., and Benson, S. K. (2017). Explicit instruction: Historical and contemporary contexts. *Learning Disabilities Research amp; Practice*, 32(3):140–148.

Iyer, V., Malik, B., Zhu, W., Stepachev, P., Chen, P., Haddow, B., and Birch, A. (2024). Exploring very low-resource translation with LLMs: The University of Edinburgh‘s submission to AmericasNLP 2024 translation task. In Mager, M., Ebrahimi, A., Rijhwani, S., Oncevay, A., Chiruzzo, L., Pugh, R., and von der Wense, K., editors, *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, pages 209–220, Mexico City, Mexico. Association for Computational Linguistics.

Jacob, J. and Grimes, C. (2003). *Kamus pengantar bahasa Kupang*. Paradigma series B / Centre for Regional Studies, UKAW. Artha Wacana Press.

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., de las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., Lavaud, L. R., Lachaux,

UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

M.-A., Stock, P., Scao, T. L., Lavril, T., Wang, T., Lacroix, T., and Sayed, W. E. (2023). Mistral 7b.

Jiao, W., Wang, W., tse Huang, J., Wang, X., Shi, S., and Tu, Z. (2023). Is chatgpt a good translator? yes with gpt-4 as the engine.

Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T. (2017). Bag of tricks for efficient text classification. In Lapata, M., Blunsom, P., and Koller, A., editors, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431, Valencia, Spain. Association for Computational Linguistics.

Kamholz, D., Pool, J., and Colowick, S. (2014). PanLex: Building a resource for panlingual lexical translation. In Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3145–3150, Reykjavik, Iceland. European Language Resources Association (ELRA).

Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press.

Kudugunta, S., Caswell, I., Zhang, B., Garcia, X., Choquette-Choo, C. A., Lee, K., Xin, D., Kusupati, A., Stella, R., Bapna, A., and Firat, O. (2023). Madlad-400: A multilingual and document-level large audited dataset.

Landis, J. R. and Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 33(2):363–374.

Lee, B. W., Cho, H., and Yoo, K. M. (2024). Instruction tuning with human curriculum. In Duh, K., Gomez, H., and Bethard, S., editors, *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 1281–1309, Mexico City, Mexico. Association for Computational Linguistics.

Leng, Y., Tan, X., Qin, T., Li, X.-Y., and Liu, T.-Y. (2019). Unsupervised pivot translation for distant languages. In Korhonen, A., Traum, D., and Màrquez, L., editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 175–183, Florence, Italy. Association for Computational Linguistics.

UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

Li, J., Zhou, H., Huang, S., Cheng, S., and Chen, J. (2024). Eliciting the Translation Ability of Large Language Models via Multilingual Finetuning with Translation Instructions. *Transactions of the Association for Computational Linguistics*, 12:576–592. Place: Cambridge, MA Publisher: MIT Press.

Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Lin, P., Ji, S., Tiedemann, J., Martins, A. F. T., and Schütze, H. (2024). Mala-500: Massive language adaptation of large language models.

Lin, X. V., Mihaylov, T., Artetxe, M., Wang, T., Chen, S., Simig, D., Ott, M., Goyal, N., Bhosale, S., Du, J., Pasunuru, R., Shleifer, S., Koura, P. S., Chaudhary, V., O'Horo, B., Wang, J., Zettlemoyer, L., Kozareva, Z., Diab, M., Stoyanov, V., and Li, X. (2022). Few-shot learning with multilingual language models.

Liu, X., Wang, L., Wong, D. F., Ding, L., Chao, L. S., Shi, S., and Tu, Z. (2021). On the copying behaviors of pre-training for neural machine translation. In Zong, C., Xia, F., Li, W., and Navigli, R., editors, *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4265–4275, Online. Association for Computational Linguistics.

Lopo, J. A. and Tanone, R. (2024). Constructing and expanding low-resource and underrepresented parallel datasets for indonesian local languages.

Lyu, C., Du, Z., Xu, J., Duan, Y., Wu, M., Lynn, T., Aji, A. F., Wong, D. F., and Wang, L. (2024). A paradigm shift: The future of machine translation lies with large language models. In Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N., editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 1339–1352, Torino, Italia. ELRA and ICCL.

Ma, C., Imani, A., Ye, H., Pei, R., Asgari, E., and Schuetze, H. (2025). Taxi1500: A dataset for multilingual text classification in 1500 languages. In Chiruzzo, L., Ritter, A., and Wang, L., editors, *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 414–439, Albuquerque, New Mexico. Association for Computational Linguistics.

UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

Ma, C., ImaniGooghari, A., Ye, H., Pei, R., Asgari, E., and Schütze, H. (2024). Taxi1500: A multilingual dataset for text classification in 1500 languages.

Manhitu, Y. (2005a). Pantun Melayu Kupang oleh Yohanes Manhitu — geocities.ws. *https://www.geocities.ws/johnmanhitu2001/PANTUN_MELAYU_KUPANG_YOHANES_MANHITU.html*. [Accessed 27-01-2025].

Manhitu, Y. (2005b). Yohanes Manhitu Pung Puisi Bahasa Melayu Kupang — geocities.ws. *https://www.geocities.ws/johnmanhitu2001/PUISI_PAKE_BAHASA_MELAYU_KUPANG.html*. [Accessed 27-01-2025].

Mao, Z. and Yu, Y. (2024). Tuning LLMs with contrastive alignment instructions for machine translation in unseen, low-resource languages. In Ojha, A. K., Liu, C.-h., Vylomova, E., Pirinen, F., Abbott, J., Washington, J., Oco, N., Malykh, V., Logacheva, V., and Zhao, X., editors, *Proceedings of the Seventh Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2024)*, pages 1–25, Bangkok, Thailand. Association for Computational Linguistics.

Merx, R., Mahmudi, A., Langford, K., de Araujo, L. A., and Vylomova, E. (2024). Low-resource machine translation through retrieval-augmented LLM prompting: A study on the Mambai language. In Ojha, A. K., Ahmadi, S., Cinková, S., Fransen, T., Liu, C.-H., and McCrae, J. P., editors, *Proceedings of the 2nd Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia (EURALI) @ LREC-COLING 2024*, pages 1–11, Torino, Italia. ELRA and ICCL.

Muennighoff, N., Wang, T., Sutawika, L., Roberts, A., Biderman, S., Le Scao, T., Bari, M. S., Shen, S., Yong, Z. X., Schoelkopf, H., Tang, X., Radev, D., Aji, A. F., Almubarak, K., Albanie, S., Alyafeai, Z., Webson, A., Raff, E., and Raffel, C. (2023). Crosslingual generalization through multitask finetuning. In Rogers, A., Boyd-Graber, J., and Okazaki, N., editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15991–16111, Toronto, Canada. Association for Computational Linguistics.

Nelson, D. L. (2008). A context-based strategy for teaching vocabulary. *The English Journal*, 97(4):33–37.

Novitasari, S., Tjandra, A., Sakti, S., and Nakamura, S. (2020). Cross-lingual machine speech chain for Javanese, Sundanese, Balinese, and Bataks speech

UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

recognition and synthesis. In Beermann, D., Besacier, L., Sakti, S., and Soria, C., editors, *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 131–138, Marseille, France. European Language Resources association.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In Isabelle, P., Charniak, E., and Lin, D., editors, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Pauzan, P. (2024). Theory in second language acquisition (recognition of concepts toward krashen's second language acquisition theory for five main hypotheses). *Journal on Education*, 6(4):20876–20888.

Peng, K., Ding, L., Zhong, Q., Shen, L., Liu, X., Zhang, M., Ouyang, Y., and Tao, D. (2023). Towards making the most of ChatGPT for machine translation. In Bouamor, H., Pino, J., and Bali, K., editors, *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 5622–5633, Singapore. Association for Computational Linguistics.

Popović, M. (2015). chrF: character n-gram F-score for automatic MT evaluation. In Bojar, O., Chatterjee, R., Federmann, C., Haddow, B., Hokamp, C., Huck, M., Logacheva, V., and Pecina, P., editors, *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal. Association for Computational Linguistics.

Post, M. (2018). A call for clarity in reporting BLEU scores. In Bojar, O., Chatterjee, R., Federmann, C., Fishel, M., Graham, Y., Haddow, B., Huck, M., Yepes, A. J., Koehn, P., Monz, C., Negri, M., Névéol, A., Neves, M., Post, M., Specia, L., Turchi, M., and Verspoor, K., editors, *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Pourdamghani, N., Aldarrab, N., Ghazvininejad, M., Knight, K., and May, J. (2019). Translating translationese: A two-step approach to unsupervised machine translation. In Korhonen, A., Traum, D., and Màrquez, L., editors, *Proceedings of*

UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

the 57th Annual Meeting of the Association for Computational Linguistics, pages 3057–3062, Florence, Italy. Association for Computational Linguistics.

Rafael, A. M. D. (2019). Interferensi fonologis penutur bahasa melayu kupang ke dalam bahasa indonesia di kota kupang. *Jurnal Penelitian Humaniora*, 20(1):47–58.

Řehůřek, R. and Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta. ELRA. `http://is.muni.cz/publication/884893/en`.

Rekrut, M. D. (1996). Effective vocabulary instruction. *The High School Journal*, 80(1):66–74.

Robinson, N., Ogayo, P., Mortensen, D. R., and Neubig, G. (2023). ChatGPT MT: Competitive for high- (but not low-) resource languages. In Koehn, P., Haddow, B., Kocmi, T., and Monz, C., editors, *Proceedings of the Eighth Conference on Machine Translation*, pages 392–418, Singapore. Association for Computational Linguistics.

Rolnick, D., Ahuja, A., Schwarz, J., Lillicrap, T., and Wayne, G. (2019). Experience replay for continual learning. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Rose, S., Engel, D., Cramer, N., and Cowley, W. (2010). Automatic Keyword Extraction from Individual Documents. In *Text Mining*, pages 1–20. John Wiley & Sons, Ltd. Section: 1 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470689646.ch1.

Schwenk, H., Chaudhary, V., Sun, S., Gong, H., and Guzmán, F. (2021a). WikiMatrix: Mining 135M parallel sentences in 1620 language pairs from Wikipedia. In Merlo, P., Tiedemann, J., and Tsarfaty, R., editors, *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1351–1361, Online. Association for Computational Linguistics.

UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

Schwenk, H., Wenzek, G., Edunov, S., Grave, E., Joulin, A., and Fan, A. (2021b). CCMatrix: Mining billions of high-quality parallel sentences on the web. In Zong, C., Xia, F., Li, W., and Navigli, R., editors, *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)*, pages 6490–6500, Online. Association for Computational Linguistics.

Sharma, P. and Li, Y. (2019). Self-supervised contextual keyword and keyphrase retrieval with self-labelling. *Preprints*.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, pages 223–231, Cambridge, Massachusetts, USA. Association for Machine Translation in the Americas.

Station, J. F. (1980–2014). Jakarta field station. Item in collection "MPI EVA corpora". The Language Archive. *https://hdl.handle.net/1839/3a60a94c-0b69-4513-bf28-de0dbea23471*.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS. In Calzolari, N., Choukri, K., Declerck, T., Doğan, M. U., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2214–2218, Istanbul, Turkey. European Language Resources Association (ELRA).

Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., Bikel, D., Blecher, L., Ferrer, C. C., Chen, M., Cucurull, G., Esiobu, D., Fernandes, J., Fu, J., Fu, W., Fuller, B., Gao, C., Goswami, V., Goyal, N., Hartshorn, A., Hosseini, S., Hou, R., Inan, H., Kardas, M., Kerkez, V., Khabsa, M., Kloumann, I., Korenev, A., Koura, P. S., Lachaux, M.-A., Lavril, T., Lee, J., Liskovich, D., Lu, Y., Mao, Y., Martinet, X., Mihaylov, T., Mishra, P., Molybog, I., Nie, Y., Poulton, A., Reizenstein, J., Rungta, R., Saladi,

UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

K., Schelten, A., Silva, R., Smith, E. M., Subramanian, R., Tan, X. E., Tang, B., Taylor, R., Williams, A., Kuan, J. X., Xu, P., Yan, Z., Zarov, I., Zhang, Y., Fan, A., Kambadur, M., Narang, S., Rodriguez, A., Stojnic, R., Edunov, S., and Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., and Purwarianti, A. (2020). IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. In Wong, K.-F., Knight, K., and Wu, H., editors, *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 843–857, Suzhou, China. Association for Computational Linguistics.

Winata, G. I., Aji, A. F., Cahyawijaya, S., Mahendra, R., Koto, F., Romadhony, A., Kurniawan, K., Moeljadi, D., Prasojo, R. E., Fung, P., Baldwin, T., Lau, J. H., Sennrich, R., and Ruder, S. (2023). NusaX: Multilingual parallel sentiment dataset for 10 Indonesian local languages. In Vlachos, A. and Augenstein, I., editors, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 815–834, Dubrovnik, Croatia. Association for Computational Linguistics.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., and Raffel, C. (2021). mT5: A massively multilingual pre-trained text-to-

UNIVERSITAS GADJAH MADA
Pendekatan Instruksional Linguistik Untuk Pelatihan Model Terjemahan Berbasis Continual Instruction Tuning Dalam Bahasa Melayu Kupang
Joanito Agili Lopo, Yunita Sari, S.Kom., M.Sc., Ph.D.; Dr.techn. Guntur Budi Herwanto, S.Kom., M.Cs.
Universitas Gadjah Mada, 2025 | Diunduh dari http://etd.repository.ugm.ac.id/

text transformer. In Toutanova, K., Rumshisky, A., Zettlemoyer, L., Hakkani-Tur, D., Beltagy, I., Bethard, S., Cotterell, R., Chakraborty, T., and Zhou, Y., editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.

Zahriani, Z. (2014). Kontektualisasi direct instruction dalam pembelajaran sains. *Lantanida Journal*, 2(1):95.

Zhang, K., Choi, Y., Song, Z., He, T., Wang, W. Y., and Li, L. (2024a). Hire a linguist!: Learning endangered languages in LLMs with in-context linguistic descriptions. In Ku, L.-W., Martins, A., and Srikumar, V., editors, *Findings of the Association for Computational Linguistics: ACL 2024*, pages 15654–15669, Bangkok, Thailand. Association for Computational Linguistics.

Zhang, W., Chan, H. P., Zhao, Y., Aljunied, M., Wang, J., Liu, C., Deng, Y., Hu, Z., Xu, W., Chia, Y. K., Li, X., and Bing, L. (2024b). Seallms 3: Open foundation and chat multilingual large language models for southeast asian languages.

Zhuang, L., Wayne, L., Ya, S., and Jun, Z. (2021). A robustly optimized BERT pre-training approach with post-training. In Li, S., Sun, M., Liu, Y., Wu, H., Liu, K., Che, W., He, S., and Rao, G., editors, *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 1218–1227, Huhhot, China. Chinese Information Processing Society of China.

Zoph, B., Yuret, D., May, J., and Knight, K. (2016). Transfer learning for low-resource neural machine translation. In Su, J., Duh, K., and Carreras, X., editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1568–1575, Austin, Texas. Association for Computational Linguistics.