

Adam Nohejl – Publications

Peer-Reviewed Journals

1. Adam Nohejl, Akio Hayakawa, Yusuke Ide, and Taro Watanabe. 2025. A Japanese dataset and efficient multilingual LLM-based methods for lexical simplification and lexical complexity prediction. *Journal of Natural Language Processing* (conditionally accepted).

Peer-Reviewed International Conference and Workshop Proceedings

1. Yusuke Ide, Joshua Tanner, Adam Nohejl, Jacob Hoffman, Justin Vasselli, Hidetaka Kamigaito, and Taro Watanabe. 2025. [CoAM: Corpus of all-type multiword expressions](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (to appear).
2. Genta Indra Winata, Frederikus Hudi, Patrick Amadeus Irawan, David Anugraha, Rifki Afina Putri, Wang Yutong, Adam Nohejl, Ubaidillah Ariq Prathama, Nedjma Ousidhoum, and 42 additional authors. 2025. [WorldCuisines: A massive-scale benchmark for multilingual and multicultural visual question answering on global cuisines](#). In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 3242–3264, Albuquerque, New Mexico. Association for Computational Linguistics. **Best Theme Paper Award**.
3. Adam Nohejl, Frederikus Hudi, Eunike Andriani Kardinata, Shintaro Ozaki, Maria Angelica Riera Machin, Hongyu Sun, Justin Vasselli, and Taro Watanabe. 2025. [Beyond film subtitles: Is YouTube the best approximation of spoken vocabulary?](#) In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 9566–9585, Abu Dhabi, UAE. Association for Computational Linguistics.
4. Justin Vasselli, Adam Nohejl, and Taro Watanabe. 2025. [Measuring the robustness of reference-free dialogue evaluation systems](#). In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 4958–4972, Abu Dhabi, UAE. Association for Computational Linguistics.
5. Matthew Shardlow, Fernando Alva-Manchego, Riza Batista-Navarro, Stefan Bott, Saul Calderon Ramirez, Rémi Cardon, Thomas François, Akio Hayakawa, Andrea Horbach, Anna Hülsing, Yusuke Ide, Joseph Marvin Imperial, Adam Nohejl, Kai North, Laura Occhipinti, Nelson Pérez Rojas, Nishat Raihan, Tharindu Ranasinghe, Martin Solis Salazar, Sanja Štajner, Marcos Zampieri, and Horacio Saggion. 2024. [The BEA 2024 shared task on the multilingual lexical simplification pipeline](#). In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*, pages 571–589, Mexico City, Mexico. Association for Computational Linguistics.
6. Yusuke Sakai, Adam Nohejl, Jiangnan Hang, Hidetaka Kamigaito, and Taro Watanabe. 2024. [Toward the evaluation of large language models considering score variance across instruction templates](#). In *Proceedings of the 7th BlackboxNLP Workshop: Analyzing and Interpreting Neural Networks for NLP*, pages 499–529, Miami, Florida, US. Association for Computational Linguistics.
7. Matthew Shardlow, Fernando Alva-Manchego, Riza Batista-Navarro, Stefan Bott, Saul Calderon Ramirez, Rémi Cardon, Thomas François, Akio Hayakawa, Andrea Horbach, Anna Hülsing, Yusuke Ide, Joseph Marvin Imperial, Adam Nohejl, Kai North, Laura Occhipinti, Nelson Pérez Rojas, Nishat Raihan, Tharindu Ranasinghe, Martin Solis Salazar, Marcos Zampieri, and Horacio Saggion. 2024. [An extensible massively multilingual lexical simplification pipeline dataset using the MultiLS framework](#). In *Proceedings of the 3rd Workshop on Tools and Resources for People with READING Difficulties (READI) @ LREC-COLING 2024*, pages 38–46, Torino, Italia. ELRA and ICCL.
8. Adam Nohejl, Akio Hayakawa, Yusuke Ide, and Taro Watanabe. 2024. [Difficult for whom? A study of Japanese lexical complexity](#). In *Proceedings of the Third Workshop on Text Simplification, Accessibility and Readability (TSAR 2024)*, pages 69–81, Miami, Florida, USA. Association for Computational Linguistics.
9. Yusuke Ide, Masato Mita, Adam Nohejl, Hiroki Ouchi, and Taro Watanabe. 2023. [Japanese lexical complexity for non-native readers: A new dataset](#). In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 477–487, Toronto, Canada. Association for Computational Linguistics.
10. Justin Vasselli, Christopher Vasselli, Adam Nohejl, and Taro Watanabe. 2023. [NAISTeacher: A prompt and rerank approach to generating teacher utterances in educational dialogues](#). In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 772–784, Toronto, Canada. Association for Computational Linguistics. **Shared Task Winner**.

Non-Refereed Proceedings

1. Adam Nohejl and Taro Watanabe. 2025. [Dispersion measures as predictors of lexical decision time, word familiarity, and lexical complexity](#). In *言語処理学会第 31 回年次大会発表論文集 (Proceedings of the 31st Annual Meeting of the Association for Natural Language Processing)*, Nagasaki, Japan.
2. Adam Nohejl. 2024. [YouTube as a vast multilingual speech corpus](#). In *Lightning Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, pages 8–14, Helsinki, Finland. Fly for Points.
3. 坂井 優介, ノヘイル アダム, 上垣外 英剛, and 渡辺 太郎. 2023. [大規模言語モデルの統一評価に向けた指示テンプレートの提案及びその評価結果の考察 \(Proposal of instruction templates for unified evaluation of LLMs and discussion of evaluation results\)](#). In 第 258 回自然言語処理研究発表会 (The 258th Natural Language Processing Research Presentation Meeting, “NL 研”). 情報処理学会 (Information Processing Society of Japan). 優秀研究賞 (Excellent Research Award).