

Marzena Karpinska

Curriculum Vitae

☎ (+1) 413-316-4261
✉ marzena.karpinska@gmail.com
🌐 <https://marzenakrp.github.io/>
in [marzena-karpinska-93049487](https://scholar.google.com/citations?user=marzena-karpinska-93049487)
G Google Scholar

Marzena Karpinska is currently a Senior Researcher at Microsoft Translator. Prior to that, she was a postdoctoral researcher at the **Manning College of Information & Computer Sciences, University of Massachusetts Amherst**, advised by **Mohit Iyyer**. She received her Ph.D. from the **Department of Language and Information Sciences at the University of Tokyo**. Her research interests lie at the intersection of natural language processing (NLP) and human-centered computing.

Education

- 2020–2024 **Postdoc:** The University of Massachusetts Amherst (USA), Manning College of Information & Computer Sciences, Natural Language Processing
Focus: *Long-form Text Processing and Generation, Machine Translation, Multilingual NLP, Summarization.*
- 2014–2020 **Ph.D.:** The University of Tokyo (Japan), Language and Information Sciences Dept.
- 2012–2014 **M.A.:** The University of Tokyo (Japan), Language and Information Sciences Dept.
- 2011–2012 **Research Stay:** The University of Tokyo (Japan), Language and Information Sciences Dept.
- 2007–2010 **M.A.:** The University of Warsaw (Poland), Oriental Studies Dept., Japanese Linguistics
- 2008–2009 **Research Stay:** Tokyo University of Foreign Studies (Japan), Japanese Studies
- 2004–2007 **B.A.:** The University of Warsaw (Poland), Oriental Studies Dept., Japanese Linguistics

Research Experience

- 2024– **Senior Researcher:** Microsoft Translator (USA), (*full-time position*)
Focus: Document-level machine translation and evaluation
- 2020–2024 **Postdoctoral Researcher:** University of Massachusetts Amherst (USA), (*full-time position*)
Focus: Document-level machine translation
Long-form generation of creative text
Multilingual question answering with focus on low-resource languages
Book-length summarization and claim verification
- 2018–2019 **Research Assistant:** The University of Tokyo (Japan), A joint research project of The University of Tokyo, Tokyo Institute of Technology, and Konan University, (*part-time position*)
Focus: Cleaning and processing multilingual corpora
Training word2vec models and aligning embeddings with MUSE
Curating cognates dataset for Romance, Germanic, and Slavic languages

Publications

Acceptance rate mostly 20%–30%

1. **[COLING-Industry]** Taishi Nakamura, Mayank Mishra, Simone Tedeschi, Yekun Chai, (...) **Marzena Karpinska**, (...), Huu Nguyen, Sampo Pyysalo (2025). “Aurora-M: Open Source Continual Pre-training for Multilingual Language and Code.” ([URL](#))
2. **[EMNLP]** **Marzena Karpinska**, Katherine Thai, Kyle Lo, Tanya Goyal, Mohit Iyyer (2024). “One Thousand and One Pairs: A “novel” challenge for long-context language models”. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)* ([URL](#))
3. **[WMT]** Tom Kocmi, Vilém Zouhar, Eleftherios Avramidis, Roman Grundkiewicz, **Marzena Karpinska**, Maja Popović, Mrinmaya Sachan, Mariya Shmatova (2024). “Error Span Annotation: A Balanced Approach for Human Evaluation of Machine Translation”. *Proceedings of the Ninth Conference on Machine Translation (WMT)* ([URL](#))
4. **[WMT Report]** Tom Kocmi, Eleftherios Avramidis, Rachel Bawden, Ondrej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, Barry Haddow, **Marzena Karpinska**, Philipp Koehn, Benjamin Marie, Kenton Murray, Masaaki Nagata, Martin Popel, Maja Popovic, Mariya Shmatova, Steinþór Steingrímsson, Vilém Zouhar (2024). “Findings of the WMT24 General Machine Translation Shared Task: The LLM Era Is Here but MT Is Not Solved Yet”. *Proceedings of the Ninth Conference on Machine Translation (WMT)* ([URL](#))
5. **[COLM]** Yekyung Kim, Yapei Chang, **Marzena Karpinska**, Aparna Garimella, Varun Manjunatha, Kyle Lo, Tanya Goyal, Mohit Iyyer (2024). “FABLES: Evaluating faithfulness and content selection in book-length summarization”. *Conference on Language Modeling (COLM)* ([URL](#))
6. **[LREC-COLING]** Anna Rogers, **Marzena Karpinska**, Ankita Gupta, Vladislav Lialin, Gregory Smelkov, Anna Rumshisky (2024). “NarrativeTime: Dense Temporal Annotation on a Timeline”. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)* ([URL](#))
7. **[WMT]** **Marzena Karpinska** and Mohit Iyyer (2023). “Large language models effectively leverage document-level context for literary translation, but critical errors persist”. *Proceedings of the Eighth Conference on Machine Translation (WMT)*, ([URL](#))
8. **[NeurIPS]** Kalpesh Krishna, Yixiao Song, **Marzena Karpinska**, John Wieting, Mohit Iyyer (2023). “Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense”. *Advances in Neural Information Processing Systems 36 (NeurIPS)*, ([URL](#))
9. **[ACL Report]** Anna Rogers, **Marzena Karpinska**, Jordan Boyd-Graber, Naoaki Okazaki (2023). “Program Chairs’ Report on Peer Review at ACL 2023”. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, xl–lxxv. Toronto, Canada: Association for Computational Linguistics. ([URL](#))
10. **[EACL Findings]** Ankita Gupta, **Marzena Karpinska**, Wenlong Zhao, Kalpesh Krishna, Jack Merullo, Luke Yeh, Mohit Iyyer, Brendan O’Connor (2023). “ezCoref: Towards Unifying Annotation Guidelines for Coreference Resolution”. *Findings of the Association for Computational Linguistics: EACL 2023*. Dubrovnik, Croatia: Association for Computational Linguistics, pp. 312–330. ([URL](#))
11. **[EMNLP]** Katherine Thai[★], **Marzena Karpinska**[★], Kalpesh Krishna, William Ray, Moira Inghilleri, John Wieting, and Mohit Iyyer (2022). “Exploring Document-Level Literary Machine Translation with Parallel Paragraphs from World Literature”. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, pp. 9882–9902. ([URL](#)) *equal contributions*[★]
12. **[EMNLP]** **Marzena Karpinska**, Nishant Raj, Katherine Thai, Yixiao Song, Ankita Gupta, and Mohit Iyyer (2022). “DEMETR: Diagnosing Evaluation Metrics for Translation”. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Abu Dhabi, United Arab

- Emirates: Association for Computational Linguistics, pp. 540–9561. ([URL](#))
13. **[COLING]** Yoshifumi Kawasaki, Maëlys Salingre, **Marzena Karpinska**, Hiroya Takamura, and Ryo Nagata (2022). “Revisiting Statistical Laws of Semantic Shift in Romance Cognates”. *Proceedings of the 29th International Conference on Computational Linguistics*. Gyeongju, Republic of Korea: International Committee on Computational Linguistics, pp. 141–151. ([URL](#))
 14. **[ANLP]** Yoshifumi Kawasaki, Maëlys Salingre, **Marzena Karpinska**, Hiroya Takamura, and Ryo Nagata (2022). “Analysis of Semantic Change in Romance Cognate Verbs Using Word Embeddings”. *Proceedings of the 28th Annual Conference of the Association for Natural Language Processing*. Hamamatsu, Japan: The Association for Natural Language Processing. ([URL](#)) *special committee award for novelty*
 15. **[EMNLP]** **Marzena Karpinska**, Nader Akoury, and Mohit Iyyer (2021). “The Perils of Using Mechanical Turk to Evaluate Open-Ended Text Generation”. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, pp. 1265–1285. ([URL](#))
 16. **[ICPhS]** **Marzena Karpinska** (2019). “How accented do Caucasian-looking vs. Asian-looking native speakers sound to a Japanese listener?” *Proceedings of the 19th International Congress of Phonetic Sciences*. Melbourne, Australia, pp 3691–3695. ([URL](#))
 17. **[Book Chapter]** **Marzena Karpinska**, Paula Kurzawska, and Katarzyna Rozanska (2019). “Digital Lingua Franca or a Culture-Specific Product Leading to Misunderstandings?” *Emoticons, Kaomoji, and Emoji: The Transformation of Communication in the Digital Age (Routledge Research in Language and Communication)* ([URL](#))
 18. **[KJEE]** Kimie Yamamura, Ryo Gakutani, **Marzena Karpinska**, Tetsuro Tanojiri, and Tom Gally (2019). “The Discourse of Kyōyō and English Education in Japan” *Komaba Journal of English Education* ([URL](#))
 19. **[RelNLP]** **Marzena Karpinska**, Bofang Li, Anna Rogers, and Aleksandr Drozd (2018). “Subcharacter Information in Japanese Embeddings: When Is It Worth It?” *Proceedings of the Workshop on the Relevance of Linguistic Structure in Neural Architectures for NLP*. Melbourne, Australia: Association for Computational Linguistics, pp. 28–37. ([URL](#))
 20. **[ICPhS]** **Marzena Karpinska**, Shodai Uchida, and Isabelle Grenon (2015). “Vowel perception by listeners from different English dialects” *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, Scotland ([URL](#))

Preprints

1. **[Preprint]** Jenna Russell, **Marzena Karpinska**, Mohit Iyyer (2025). “People who frequently use ChatGPT for writing tasks are accurate and robust detectors of AI-generated text”. ARXIV: 2501.15654 [cs.CL] (under review)
2. **[Preprint]** Yekyung Kim, Jenna Russell, **Marzena Karpinska**, Mohit Iyyer (2025). “One ruler to measure them all: Benchmarking multilingual long-context language models”. ARXIV: 2503.01996 [cs.CL]
3. **[Preprint]** Abhinav Kumar, Jaechul Roh, Ali Naseh, **Marzena Karpinska**, Mohit Iyyer, Amir Houmansadr, Eugene Bagdasarian (2025). “OverThink: Slowdown Attacks on Reasoning LLMs”. ARXIV: 2502.02542 [cs.CL] (under review)
4. **[Preprint]** Shane Arora★, **Marzena Karpinska**★, Hung-Ting Chen, Ipsita Bhattacharjee, Mohit Iyyer, Eunsol Choi (2024). “CaLMQA: Exploring culturally specific long-form question answering across 23 languages”. ARXIV: 2406.17761 [cs.CL] *equal contributions*★ (under review)
5. **[Preprint]** Tom Kocmi, Eleftherios Avramidis, Rachel Bawden, Ondrej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, Barry Haddow, **Marzena Karpinska**, Philipp Koehn, Benjamin Marie, Kenton Murray, Masaaki Nagata,

Martin Popel, Maja Popovic, Mariya Shmatova, Steinþór Steingrímsson, Vilém Zouhar (2024).
“Preliminary WMT24 Ranking of General MT Systems and LLMs”. ARXIV: 2407.19884 [cs.CL]

Ongoing Projects

1. **LitMT:** Collecting readers’ feedback on machine translated literature in order to further improve machine translation systems and train passage-level evaluation metrics ([URL](#)).
2. **Human-Computer Interaction:** Investigating how machine translation systems can be introduced in the traditional translation pipeline.
3. **NoCha:** Evaluating the ability of long-context language models in verifying factual claims within book-length texts ([URL](#)).

Teaching Experience

- 2024 **Guest Lecture:** University of Massachusetts Amherst (USA)
Human-centered Machine Translation (COMP-LIT 481)
- 2023 **Guest Lecture:** University of Massachusetts Amherst (USA)
Machine Translation on Paragraph-level (CS 685)
- 2022 **Guest Lecture:** University of Massachusetts Amherst (USA)
Biases in Large Language Models (PHILL 755)
- 2018–2019 **Instructor:** Tama Art University (Japan), (*part-time position*)
- 2014–2018 **Instructor:** Nippon Institute of Technology (Japan), (*part-time position*)
- 2015–2018 **Teaching Assistant:** The University of Tokyo (Japan), (*part-time position*)
- 2012–2013 **Teaching Assistant:** The University of Tokyo (Japan), (*part-time position*)

Mentoring Experience

Ankita Gupta, Katherine Thai, Yekyung Kim, Jenna Russell (Ph.D. students, CS)

Nishant Raj, William Ray, Anirudh Atmakuru, Anmol Mekala (Master’s students, CS)

Jiarui Liu, Ipsita Bhattacharjee, Michael Ifejiagwa, Shreyan Mallik, Agneshka Rohra, Sofia Simonoff, Minh Le, Alisha Srivastava, Emir Korukluoglu (Undergraduate students, CS)

Invited Talks

- 08/2024 UNSW Sydney, NLP Seminar
Evaluating Factuality in Long-Context Language Models
- 07/2024 The Royal Melbourne Institute of Technology, NLP Seminar
Evaluating Factuality in Long-Context Language Models
- 07/2024 The University of Melbourne, NLP Seminar
Evaluating Factuality in Long-Context Language Models

05/2023 Instituto Superior Técnico & Unbabel Seminar

Paragraph-level Machine Translation of Literary Texts with Large Language Models

04/2023 Microsoft MT Reading Group

Paragraph-level Machine Translation of Literary Texts with Large Language Models

01/2023 Polish Academy of Sciences, NLP Seminar

Evaluation of Long-form Text Generation

Awards and Scholarships

- Competitive Research Grant for Ph.D. Students awarded based on the research proposal, 2017–2019 (granted yearly; ~¥600,000)
- Research Student Scholarship of Ministry of Education of Japan, 2008–2009 and 2011–2017 (~¥11,500,000)
- Scholarship awarded yearly based on the GPA by the Rector of the University of Warsaw, 2005–2010

Service and Outreach

Reviewing: ACL, EMNLP, EACL, ICPHS, NLP Workshops, ESSLLI

Area ARR June (EMNLP 2024; ★ recognized as an outstanding AC), ARR August 2024, ARR
Chair: October (NAACL 2024)

Other: Statistical analysis of peer-review data from the Association for Computational Linguistics (ACL) 2023

Industry Experience

2014–2020 **Language Specialist:** Innovative Language Learning (ILL) (Japan), (*part/full-time position*)

Tasks Developing curricula, assessments, and study materials for over 30 languages on the ILL digital learning platform ([URL](#))

Designing experiments to measure and improve students' performance

Utilizing NLP tools for content development

Improving user experience with the Alexa voice assistant for the ILL language skill

2008–2020 **Translator:** Japanese-Polish-English Translator and Interpreter (*freelance*)

Experience Various translations including manga, legal documents, technical documentation, oral presentations, and formal events for many high profile clients such as Deloitte, Sumitomo Corporation, Toyota, and Polish Embassy in Japan

Skills and Expertise

- *Coding:* PYTHON, R, JAVASCRIPT, CSS, L^AT_EX;
- *Libraries:* PyTorch, Scikit-Learn, Numpy, Scipy;
- *Theoretical frameworks:* distributional semantics, frame semantics, cognitive semantics, pragmatics, discourse analysis, sociolinguistics, sociophonetics, translation theory;

- *Research Skills*: designing controlled experiments, collecting participant data, performing statistical analysis (R), ethics of data collection;
- *Languages*: English, Polish, Japanese (native or near-native), German (intermediate), Chinese, Czech, French, Russian (beginner), broad knowledge of typological traits of over 30 languages available on the ILL platform ([URL](#))