Michael Hahn

Contact

Saarland University
Department of Language Science and Technology
Saarbrücken, Germany
Website: https://www.mhahn.info/
E-mail: mhahn@lst.uni-saarland.de

Employment

2022–present **Saarland University**, Saarbrücken, Germany
Professor (tenure track, W2)
Chair for Language, Computation, and Cognition
Department of Language Science and Technology, Faculty of Humanities
Faculty of Mathematics and Computer Science (co-opted)

2018 **Facebook AI Research**, Paris, France
Research Intern. Mentor: Marco Baroni

Education

2016–2022 **Stanford University**
Ph.D. in Linguistics
Dissertation: *Computational and Communicative Efficiency in Language*
Committee: Judith Degen, Dan Jurafsky (co-advisors), Michael C. Frank

2019 **Massachusetts Institute of Technology**
Department of Brain and Cognitive Sciences
Visiting PhD student. Advisor: Edward Gibson

2015–2016 **University of Edinburgh**
School of Informatics
Postgraduate Research Visiting Student. Advisor: Frank Keller

2014–2015 **University of Tübingen**
M.Sc. in Mathematics

2014 26th European Summer School in Logic, Language and Information

2011-2014 **University of Tübingen**
B.Sc. in Mathematics

2010-2013 **University of Tübingen**
B.A. in Computational Linguistics

Fellowships and Awards

2024 Best Paper Award, ACL 2024 *(awarded for Hahn & Rofin, "Why are Sensitive Functions Hard for Transformers?").*

2024 SAC Award: Machine Learning for NLP, ACL 2024 *(awarded for Hahn & Rofin, "Why are Sensitive Functions Hard for Transformers?").*

2024 Outstanding Area Chair, EMNLP 2024.

2024 Second Prize, Mechanistic Interpretability Workshop at ICML 2024 *(awarded for Huang et al, "InversionView: A General-Purpose Method for Reading Information from Neural Activations", final version published at NeurIPS 2024).*

2022 Sayan Gul Award for Best Undergraduate Paper, CogSci 2022 *(awarded for Rathi, Hahn, Futrell, "Explaining patterns of fusion in morphological paradigms using the memory–surprisal tradeoff").*

2021 Best Paper Award, SIGTYP 2021 *(awarded for Rathi, Hahn, Futrell, "Information-Theoretic Characterization of Morphological Fusion", final version published at EMNLP 2021).*

2019–2022 Stanford Interdisplicinary Graduate Fellowship *(competetive fellowship for Stanford Ph.D. students with interdisciplinary research project).*

2016–2019 Stanford University PhD Fellowship.

2017 CUNY Student Travel Grant.

2011–2016 German National Academic Foundation.
(*Studienstiftung des Deutschen Volkes*)

Publications

Refereed Journal Articles

2024 **Michael Hahn** and Xue-Xin Wei. A unifying theory explains seemingly contradictory biases in perceptual estimation. *Nature Neuroscience*. 27:793–804.

2023 **Michael Hahn** and Frank Keller. Modeling Task Effects in Human Reading with Neural Network-based Attention. *Cognition*. 230:105289.

2023 Thomas Hikaru Clark, Clara Meister, Tiago Pimentel, **Michael Hahn**, Ryan Cotterell, Richard Futrell and Roger Levy. A Cross-Linguistic Pressure for Uniform Information Density in Word Order. *Transactions of the Association for Computational Linguistics (TACL).*

2022 **Michael Hahn**, Richard Futrell, Roger Levy, and Edward Gibson. A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences of the United States of America (PNAS).* 119(43):e2122602119.

2022 **Michael Hahn** and Yang Xu. Crosslinguistic word order variation reflects evolutionary pressures of dependency and information locality. *Proceedings of the National Academy of Sciences of the United States of America (PNAS).* 119(24):e2122604119.

2022 Richard Futrell and **Michael Hahn**. Information theory as a bridge between language function and language form. *Frontiers in Communication*. 7:657725.

2022 **Michael Hahn**, Rebecca Mathew, and Judith Degen. Morpheme ordering across languages reflects optimization for processing efficiency. *Open Mind: Discoveries in Cognitive Science*. 5:208–232.

2021 **Michael Hahn**, Dan Jurafsky, and Richard Futrell. Sensitivity as a complexity measure for sequence classification tasks. *Transactions of the Association for Computational Linguistics (TACL)*, 9:891–908.

2021 **Michael Hahn**, Judith Degen, and Richard Futrell. Modeling word and morpheme order in natural language as an efficient tradeoff of memory and surprisal. *Psychological Review*, 128(4):726–756.

2020 **Michael Hahn**, Dan Jurafsky, and Richard Futrell. Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 117(5):2347–2353.

2020 **Michael Hahn**. Theoretical limitations of self-attention in neural sequence models. *Transactions of the Association for Computational Linguistics (TACL)*, 8:156–171.

2019 **Michael Hahn** and Marco Baroni. Tabula nearly rasa: Probing the linguistic knowledge of character-level neural language models trained on unsegmented text. *Transactions of the Association for Computational Linguistics (TACL)*, 7:467–484.

2019 **Michael Hahn** and Richard Futrell. Estimating predictive rate-distortion curves via neural variational inference. *Entropy*, 21(7):640.

2015 **Michael Hahn** and Frank Richter. Henkin Semantics for reasoning with natural language. *Journal of Language Modeling*, 3(2):513–568.

Papers in Refereed Conference Proceedings

2025 Xinting Huang, Andy Yang, Satwik Bhattamishra, Yash Sarrof, Andreas Krebs, Hattie Zhou, Preetum Nakkiran, and **Michael Hahn**. A Formal Framework for Understanding Length Generalization in Transformers. *International Conference on Learning Representations (ICLR)*.

2024 Xinting Huang, Madhur Panwar, Navin Goyal, and **Michael Hahn**. InversionView: A General-Purpose Method for Reading Information from Neural Activations. *Advances in Neural Information Processing Systems (NeurIPS)*.

2024 Yash Sarrof, Yana Veitsman, and **Michael Hahn**. The Expressive Capacity of State Space Models: A Formal Language Perspective. *Advances in Neural Information Processing Systems (NeurIPS)*.

2024 Satwik Bhattamishra, **Michael Hahn**, Phil Blunsom, and Varun Kanade. Separations in the Representational Capabilities of Transformers and Recurrent Architectures. *Advances in Neural Information Processing Systems (NeurIPS)*.

2024 **Michael Hahn** and Mark Rofin. Why are Sensitive Functions Hard for Transformers? In *Proceedings of the 2024 Annual Conference of the Association for Computational Linguistics (ACL)*. **Award: Best Paper Award.**

2024 Kate McCurdy and **Michael Hahn**. Lossy Context Surprisal Predicts Task-Dependent Patterns in Relative Clause Processing. In *Proceedings of the 28th Conference on Computational Natural Language Learning (CoNLL)*.

2024 Siyu Tao, Lucia Donatelli, and **Michael Hahn**. More frequent verbs are associated with more diverse valency frames: Efficient language design at the lexicon-grammar interface. In *Proceedings of the 2024 Annual Conference of the Association for Computational Linguistics (ACL)*.

2024 Hailin Hao, Yang Yang, and **Michael Hahn**. Information Locality in the Processing of Classifier-Noun Dependencies in Mandarin Chinese. In *Proceedings of the 46th Annual Meeting of the Cognitive Science Society (CogSci)*.

2023 Hailin Hao, **Michael Hahn**, and Elsi Kaiser. How do syntactic statistics and semantic plausibility modulate local coherence effects. In *Proceedings of the 45th Annual Meeting of the Cognitive Science Society (CogSci)*.

2022 Songpeng Yan, **Michael Hahn**, and Frank Keller. Modeling fixation behavior in reading with character-level neural attention. In *Proceedings of the 44th Annual Meeting of the Cognitive Science Society (CogSci)*.

2022 Neil Rathi, **Michael Hahn**, and Richard Futrell. Explaining patterns of fusion in morphological paradigms using the memory–surprisal tradeoff. In *Proceedings of the 44th Annual Meeting of the Cognitive Science Society (CogSci)*. **Award: Sayan Gul Award for Best Undergraduate Paper.**

2021 Neil Rathi, **Michael Hahn***, and Richard Futrell*. An information-theoretic characterization of morphological fusion. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*, 10115–10120.

2020 John Hewitt, **Michael Hahn**, Surya Ganguli, Percy Liang, and Christopher D. Manning. RNNs can generate bounded hierarchical languages with optimal memory. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*, 1978–2010.

2019 **Michael Hahn**, Frank Keller, Yonatan Bisk, and Yonatan Belinkov. Character-based surprisal as a model of human reading in the presence of errors. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society (CogSci)*, 401–407.

2018 **Michael Hahn**, Judith Degen, Noah Goodman, Dan Jurafsky, and Richard Futrell. An information-theoretic explanation of adjective ordering preferences. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society (CogSci)*, 1766–1771.

2018 **Michael Hahn**, Andreas Krebs, and Howard Straubing. Wreath products of distributive forest algebras. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2018)*, 512–520.

2016 **Michael Hahn** and Frank Keller. Modeling human reading with neural attention. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016)*, 85–95.

2015 **Michael Hahn**, Andreas Krebs, Klaus-Jörn Lange, and Michael Ludwig. Visibly counter languages and the structure of NC^1. In *Mathematical Foundations of Computer Science 2015 – 40th International Symposium, MFCS 2015*, 384–394.

2014 **Michael Hahn**. Predication and NP structure in an omnipredicative language: The case of Khoekhoe. In *Proceedings of the 21st International Conference on Head-Driven Phrase Structure Grammar*, 238–258. CSLI Publications.

2013 **Michael Hahn**. Word order variation in Khoekhoe. In *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar*, 48–68. Stanford, CSLI Publications.

2013 Niels Ott, Ramon Ziai, **Michael Hahn**, and Detmar Meurers. CoMeT: Integrating different levels of linguistic modeling for meaning assessment. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval)*, 608–616.

2012 **Michael Hahn**. Arabic relativization patterns: A unified HPSG analysis. In *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar*, 144–164. Stanford, CSLI Publications.

2012 **Michael Hahn** and Detmar Meurers. Evaluating the meaning of answers to reading comprehension questions: A semantics-based approach. In *Proceedings of the 7th Workshop on Innovative Use of NLP for Building Educational Applications (BEA7)*, 326–336. Association for Computational Linguistics.

2011 **Michael Hahn** and Detmar Meurers. On deriving semantic representations from dependencies: A practical approach for evaluating meaning in learner corpora. In *Proceedings of the Int. Conference on Dependency Linguistics (Depling 2011)*, 94–103.

2011 **Michael Hahn**. Null conjuncts and bound pronouns in Arabic. In *Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar*, 60–80. Stanford, CSLI Publications.

Contributions to edited volumes

2014 **Michael Hahn** and Detmar Meurers. On deriving semantic representations from dependencies: A practical approach for evaluating meaning in learner corpora. In Kim Gerdes, Eva Hajicová, and Leo Wanner, editors, *Dependency Theory*, Frontiers in AI and Applications Series, 94–103. IOS Press, 2014.

Invited Talks

2025 *A Formal Framework for Understanding Length Generalization in Transformers*, StefSM lab at Chalmers University of Technology, 2025, Sweden, May 2025.

2025 *Why are Sensitive Functions Hard for Transformers?*, Invited Plenary at Information Theory and Applications (ITA) 2025, San Diego, CA, US, February 2025.

2024 *Understanding Language Models via Theory and Interpretability*, Lightning talk at DFKI Directors' Retreat 2024, Frankfurt, Germany, December 2024.

2024 *Understanding Language Models via Theory and Interpretability*, NLP/Text-as-Data Speaker Series, New York University, NY, US, December 2024.

2024 *Understanding Language Models via Theory, Interpretability, and Humans*, Computational Linguistics and Linguistic Theory, Universitat Pompeu Fabra, Barcelona, Spain, October 2024.

2024 *Understanding Language Models via Theory, Interpretability, and Humans*, ILCC/CDT NLP seminar, University of Edinburgh, UK, September 2024.

2024 *A Model of Language Processing as Resource-Rational Sequence Prediction*, Colloquium, Department of Linguistics, University of Potsdam, Germany, April 2024.

2023 *A Model of Language Processing as Resource-Rational Sequence Prediction*, Cognitive Science Colloquium, IIT Delhi, India, November 2023.

2023 *Modeling word and morpheme order as an efficient tradeoff of memory and surprisal*, Colloquium General Linguistics, University of Tübingen, Germany, July 2023.

2023 *A Resource-Rational Model of Human Processing of Recursive Linguistic Structure*, Colloquium English Linguistics, University of Frankfurt, Germany, May 2023.

2022 *A Resource-Rational Model of Human Processing of Recursive Linguistic Structure*, Montréal Computational & Quantitative Linguistics Lab, McGill University, Canada, November 2022.

2022 *Comprehending Language in Humans and Machines*, Wangxuan Institute of Computer Technology, Peking University, China, March 2022.

2021 *Sensitivity as a Complexity Measure for Sequence Classification Tasks*, NLP Talk Series, Microsoft Research Lab India, India, November 2021.

2021 *Cognition Constrains Linguistic Diversity in Word Order*, Computational Psycholinguistics Lab, Massachusetts Institute of Technology, US, June 2021.

2021 *Memory Efficiency Predicts Ordering Universals in Language*, Colloquium, Department of Linguistics, University of Düsseldorf, Germany, April 2021.

2021 *Sensitivity as a Complexity Measure for Sequence Classification Tasks*, Singh Lab, Department of Computer Science, UC Irvine, US, March 2021.

2021 *Word Order as an Efficient Tradeoff of Memory and Surprisal*, Colloquium, Department of Linguistics, UT Austin, US, February 2021.

2021 *An Information-Theoretic Explanation of Adjective Ordering Preferences*, TExMod2020: Theoretical and Experimental Approaches to Modification. Tübingen, Germany, January 2021.

2020 *A Neural Noisy-Channel Model of Structural Forgetting*, TedLab, Massachusetts Institute of Technology, US, November 2020.

2020 *Word Order Universals Optimize Communicative Efficiency*, Cognitive Lexicon Laboratory, University of Toronto, Canada, August 2020.

2019 *Crosslinguistic Word Orders Optimize Efficiency of Human Communication and Processing*, Harvard NLP, Harvard University, US, July 2019.

2018 *Explaining Syntactic Universals by Optimizing Grammars.* TedLab, Massachusetts Institute of Technology, US, November 2018.

Conference Presentations

2025 *Emergent Stack Representations in Modeling Counter Languages Using Transformers* (work with Aviral Gupta, Utkarsh Tiwari; presented by Aviral Gupta and Utkarsh Tiwari), ICLR World Models Workshop, April 2025.

2025 *Information Locality in the Processing of English Object Relative Clauses* (work with Hailin Hao, Weijie Xu, Richard Futrell; presented by Hailin Hao), Human Sentence Processing Conference, March 2025.

2024 *InversionView: A General-Purpose Method for Reading Information from Neural Activations* (work with X. Huang, M. Panwar, N. Goyal; presented by Xinting Huang), Mechanistic Interpretability Workshop at ICML 2024, July 2024. **Award: Second Prize.**

2024 *The Identifiability of Bayesian Models of Perceptual Decision* (work with Xue-Xin Wei; presented by Xue-Xin Wei), Vision Sciences Society Annual Meeting, May 2024.

2022 *Modeling Fixations with Neural Attention* (work with Songpeng Yan and Frank Keller), Human Sentence Processing Conference, March 2022.

2021 *An information-theoretic characterization of morphological fusion* (work with Neil Rathi and Richard Futrell, presented by Neil Rathi), SIGTYP, June 2021.

2020 *Lexical Effects in Structural Forgetting: Evidence for Experience-Based Accounts and a Neural Network Model* (work with Richard Futrell and Edward Gibson), 33rd Annual CUNY Conference on Human Sentence Processing 2020, March 2020.

2019 *Crosslinguistic word orders enable an efficient tradeoff of memory and surprisal* (work with Judith Degen, Richard Futrell), 32nd Annual CUNY Conference on Human Sentence Processing, University of Colorado Boulder, March 2019.

2019 *Testing Functional Explanations of Word Order Universals* (work with Richard Futrell), 32nd Annual CUNY Conference on Human Sentence Processing, University of Colorado Boulder, USA, March 2019.

2018 *Testing Functional Explanations of Word Order Universals.* (work with Richard Futrell) CAMP 2018, USC, Los Angeles, USA, November 2018.

2018 Poster: *Mutual Information Impacts Adjective Ordering Across Languages.* (work with Judith Degen, Dan Jurafsky, Noah Goodman, Richard Futrell), 31st Annual CUNY Conference on Human Sentence Processing, UC Davis, USA, March 2018.

2018 Poster: *Exploring Adjective Ordering Preferences via Artificial Language Learning.* (work with Judith Degen, Richard Futrell), 31st Annual CUNY Conference on Human Sentence Processing, UC Davis, USA, March 2018.

2017 *Exploring Adjective Ordering Preferences via Artificial Language Learning.* (work with Judith Degen, Richard Futrell), California Meeting on Psycholinguistics, 2017, UCLA, Los Angeles, USA, November 2017.

2017 *Modeling Task Effects in Reading with Neural Attention.* (work with Frank Keller), 30th Annual CUNY Conference on Human Sentence Processing, MIT, March 2017.

2010 *Agreement and Complex Predicates in Modern Standard Arabic*. Generative Grammatik des Nordens, Berlin, Germany, July 2010.

2009 *Nichtlokale Abhängigkeiten im Hocharabischen/Nonlocal Dependencies in Modern Standard Arabic*. Workshop on Grammar Theory and Grammar Implementation, Berlin, Germany, May 2009.

Teaching

2025 Instructor, *Large Language Models* (introductory seminar), Saarland University.

2025 Instructor, *Neural Networks in Brains and Computers* (seminar), Saarland University.

2024 Instructor, *Foundations of Mathematics* (lecture), Saarland University.

2024 Instructor, *Aligning Language Models with Human Preferences, Methods and Challenges* (seminar), Saarland University.

2024 Instructor, *Neural Networks in Brains and Computers* (seminar), Saarland University.

2023 Instructor, *Foundations of Mathematics* (lecture), Saarland University.

2020 Teaching Assistant, *Natural Language Understanding (CS 224U)*, with Christopher Potts, Stanford University.

2019 Teaching Assistant, *Natural Language Processing with Deep Learning (CS 224N)*, with Christopher D. Manning, Stanford University.

2017–2018 Linguistics Corpus Teaching Assistant, with Christopher D. Manning, Stanford University.

2018 Web-based Experimental Methods Workshop, Stanford University.

2012 Teaching Assistant, *Grammar Formalisms in Computational Linguistics*, with Detmar Meurers, University of Tübingen.

Research Advising and Mentoring

PhD Students

2023–present Xinting Huang, Saarland University.

Postdoctoral Scholars

2023–present Kate McCurdy, Saarland University.

M.Sc. Students at Saarland University

2023–present Mark Rofin

2023–present Aleksandra Bakalova

2023–present Yash Raj Sarrof

2023–present Yana Veitsman

2024–present Entang Wang

2024–present Uyen Hoang

2023–2025 Doreen Osmelak

Undergraduate Students

2024 Alireza Amiri (summer internship at Saarland University).

2020 Rebecca Mathew (CSLI Summer Internship Program, CSLI, Stanford University).

High-School Students

2020–2022 Neil Rathi (Palo Alto High School, now undergraduate at Stanford University).

2022 Adam Farris (San Mateo High School, now undergraduate at Stanford University).

Funding Awarded

2023–2026 Broad-Coverage Models of Memory and Surprisal in Human Sentence Processing Collaborative Research Center SFB 1102 - Information Density and Linguistic Encoding (IdeaL). SFB-internal grant (reviewed SFB-internally and funded from central funds). EUR 166,026.

Scientific Associations

2025 ELLIS Member, 2025.

2025 ELIZA Fellow, 2025.

Organization of Scientific Meetings

2025 7th Workshop on Research in Computational Linguistic Typology and Multilingual NLP (SIGTYP 2025).

2024 6th Workshop on Research in Computational Linguistic Typology and Multilingual NLP (SIGTYP 2024).

Reviewing

2023–present **Area Chair**, ACL Rolling Review (2024–present), CogSci (2024–present), COLM (2024–present), EMNLP (2023).

2024 **Award Committee**, EMNLP 2024.

2015–present **Conference Reviewing**, ACL Rolling Review (2022–present), ACL (2021–present), ICML (2024–present), ICLR (2024–present), NeurIPS (2023–present), AAAI (2025–present), EACL (2021–present), EMNLP (2021–present), CogSci (2020–present), CoNLL (2020–present), MFCS (2019), STACS (2015).

2019–present **Journal Reviewing (ad-hoc)**, Nature Communications (2024), TACL (2025), Cognition (2020), Open Mind (2021, 2023, 2025), Neural Networks (2021), Language (2024), Glossa: A Journal of General Linguistics (2019, 2020, 2022), Neuropsychologia (2019), Journal of Cognitive Psychology (2022), PeerJ (2019, 2022), Journal of Experimental Psychology: Learning, Memory, and Cognition (2022), Journal of Experimental Psychology: General (2025), Computational Linguistics (2024), eLife (2023, 2024), Journal of Memory and Language (2024), Communications Psychology (2024).

2024–present **Journal Reviewing (standing review committee)**, Computational Linguistics (2024–present).

2023–present **Grant Reviewing**, Dutch Research Council (NWO, 2023), German Research Foundation (DFG, 2023), Israel Science Foundation (ISF, 2024), Swiss National Science Foundation (SNSF, 2024).

Outreach

2024 *Wie lernen Computer Sprache?* (How do computers learn language?). Talk at *Kinderuni Saar* at Saarland University. *(50-minute talk at outreach activity for children aged 8-12 years.)*

2024 *Processing Language in Humans and Machines.* Talk at *Tagung der Computerlinguistik-Studierenden (TaCoS). (Invited talk at a conference of undergraduate and graduate students of Computational Linguistics in Germany.)*

Departmental Service

2024 Coordinator, ELLIS Pre-NeurIPS Fest, Saarland Informatics Campus *(initiated and organized a poster session for NeurIPS papers from Saarbrücken).*

2024 Hiring Committee Language Technology, Saarland University.

2020 CSLI Summer Internship Admissions Committee, Stanford University.

2019 QP Fest Committee, Stanford Linguistics Department.

2016–2017 Social Committee, Stanford Linguistics Department.