William Merrill

https://lambdaviking.com/
willm[æt]nyu.edu
Last updated April 22, 2022

Research Interests

Broad

NLP, deep learning, formal languages and automata, formal semantics, historical linguistics, computational complexity

Specific

- Formal capabilities and inductive biases of neural networks for language
- Theory of self-supervised learning and distributional semantics
- Robustness and interpretability of NLP models

Professional Experience

Google	2022	**Student Researcher**, automata theory
AI2	2019–2021	**Predoc. Young Investigator** with AllenNLP
Google	2018	**Software Engineering Intern** *"Exceeds expectations" rating; return offer*
Boston College	2017	**Research Intern** in Language Learning Lab
New York University	2013–2015	**Research Intern** in Morphology Lab

Education

New York University	2021–	**Ph.D.** at Center for Data Science
Yale University	2015–2019	**B.Sc.** with distinction in Computer Science **B.A.** with distinction in Linguistics Thesis: Sequential neural networks as automata *Cum laude; note of excellence on thesis*

Mentors and Group Affiliations

Tal Linzen	ML^2, NYU	2022
Noah A. Smith, Yoav Goldberg, Roy Schwartz	AllenNLP, $(AI)^2$	2019–
Robert Frank, Dana Angluin	CLAY, Yale	2016–2019
Joshua Hartshorne, Sven Dietz	L^3, Boston College	2017
Alec Marantz, Phoebe Gaston	MorphLab, NYU	2013–2015

Publications

William Merrill and Nikolaos Tsilivis. Extracting finite automata from RNNs using state merging, 2022. URL https://arxiv.org/abs/2201.12451.

William Merrill, Ashish Sabharwal, and Noah A. Smith. Saturated transformers are constant-depth threshold circuits, 2021a. URL https://arxiv.org/abs/2106.16213.

Matt Gardner, William Merrill, Jesse Dodge, Matthew E. Peters, Alexis Ross, Sameer Singh, and Noah Smith. Competency problems: On finding and removing artifacts in language data, 2021.

William Merrill, Yoav Goldberg, Roy Schwartz, and Noah A. Smith. Provable Limitations of Acquiring Meaning from Ungrounded Form: What Will Future Language Models Understand? *Transactions of the Association for Computational Linguistics*, 9:1047–1060, 09 2021b. ISSN 2307-387X. doi: 10.1162/tacl_a_00412. URL https://doi.org/10.1162/tacl_a_00412.

William Merrill, Vivek Ramanujan, Yoav Goldberg, Roy Schwartz, and Noah Smith. Effects of parameter norm growth during transformer training: Inductive bias from gradient descent, 2021c.

William Merrill, Gail Weiss, Yoav Goldberg, Roy Schwartz, Noah A. Smith, and Eran Yahav. A formal hierarchy of RNN architectures. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 443–459, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.43. URL https://www.aclweb.org/anthology/2020.acl-main.43.

Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Doug Burdick, Darrin Eide, Kathryn Funk, Yannis Katsis, Rodney Kinney, Yunyao Li, Ziyang Liu, William Merrill, Paul Mooney, Dewey Murdick, Devvret Rishi, Jerry Sheehan, Zhihong Shen, Brandon Stilson, Alex Wade, Kuansan Wang, Nancy Xin Ru Wang, Chris Wilhelm, Boya Xie, Douglas Raymond, Daniel S. Weld, Oren Etzioni, and Sebastian Kohlmeier. CORD-19: The COVID-19 open research dataset, 2020.

William Merrill. On the linguistic capacity of real-time counter automata, 2020. URL https://arxiv.org/abs/2004.06866.

William Merrill. Sequential neural networks as automata. In *Proceedings of the Workshop on Deep Learning and Formal Languages: Building Bridges*, pages 1–13, Florence, August 2019. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/W19-3901.

William Merrill, Lenny Khazan, Noah Amsel, Yiding Hao, Simon Mendelsohn, and Robert Frank. Finding hierarchical structure in neural stacks using unsupervised parsing. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 224–232, Florence, Italy, August 2019a. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/W19-4823.

William Merrill, Gigi Stark, and Robert Frank. Detecting syntactic change using a neural part-of-speech tagger. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 167–174, Florence, Italy, August 2019b. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/W19-4721.

Yiding Hao, William Merrill, Dana Angluin, Robert Frank, Noah Amsel, Andrew Benz, and Simon Mendelsohn. Context-free transductions with neural stacks. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 306–315, Brussels, Belgium, November 2018. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/W18-5433.

Jungo Kasai, Robert Frank, Pauli Xu, William Merrill, and Owen Rambow. End-to-end graph-based TAG parsing with neural networks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans,*

Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers), pages 1181–1194, 2018. URL https://aclanthology.info/papers/N18-1107/n18-1107.

William Merrill. A semantics of subordinate clauses using delayed evaluation. *Toronto Undergraduate Linguistics Conference*, 2018. URL https://ling.auf.net/lingbuzz/003487.

Invited Talks

- **MILA**, ML for Code Seminar, 2022
  Saturated Transformers are Constant-Depth Threshold Circuits
- **MIT**, CompLang Seminar, 2022
  Language Models Have Implicit Entailment Semantics
- **NYU**, Semantics Seminar, 2022
  Distributional Learnability of Entailment
- **Google**, Automata Research Team, 2022
  Neural Networks as Automata
- **ArthurAI**, Journal Club, 2021
  Competency Problems: On Finding and Removing Artifacts in Language Data
- **EMNLP**, ML Track, 2021
  Competency Problems: On Finding and Removing Artifacts in Language Data
- **EMNLP**, ML Track, 2021
  Parameter Norm Growth During Transformer Training: Inductive Bias From Gradient Descent
- **AI2**, All Hands, 2021
  Provable Limitations of Acquiring Meaning from Ungrounded Form: What Will Future Language Models Understand?
- **UW**, Noah's ARK, 2020
  Provable Limitations of Acquiring Meaning from Ungrounded Form: What Will Future Language Models Understand?

- **EMNLP**, Blackbox NLP, 2018
  Context-Free Transductions with Neural Stacks
- **Packer Collegiate Institute**, Science Symposium, 2018
  Neural networks, L2 Acquisition, and the Voynich
- **CodeHaven**, 2018
  Programming, Language, and the Book of Thoth
- **UToronto**, TULCon, 2018
  A Semantics of Subordinate Clauses Using Delayed Evalation

Teaching Assistant Experience

Yale	CPSC 477	NLP	Dragomir Radev	Spring 2019
	CPSC 477	NLP	Dragomir Radev	Spring 2018
	CPSC 470	AI	Dragomir Radev	Fall 2017

Service

Reviewing

ARR	Jan 2022	2 review
ARR	Dec 2021	3 reviews
ARR	Nov 2021	1 review
CL	2021	1 review
ACL	2021	6 reviews
EACL	2021	4 reviews
EMNLP	2020	2 reviews
Neural Networks	2020	1 review

Organization and Teaching

NYC AI School	2022	Volunteer instructor
AllenNLP Hackathon	2021	Technical support
Yale Tangut Language Workshop	2018	Workshop facilitator
Yale NACLO	2017	Student volunteer
Yale Kitan Language Workshop	2016	Workshop facilitator
CodeHaven	2016–2018	Student volunteer
Splash at Yale	2016–2017	Student instructor

Selected Public Software

- **StackNN**: Differentiable stacks, queues, and dequeues in PyTorch
- **Voynich2Vec**: Word embedding analysis of the Voynich manuscript
- **The Book of Thoth**: A puzzle game with dynamic spell casting in Ancient Egyptian
- Contributor to open-source NLP framework **AllenNLP**

Blog Posts

NLP

- Capsule networks for NLP
- Review: Learning to transduce with unbounded memory
- Word2vec analysis of the Voynich manuscript

Translations

- The Wanderer (Old English)
- After Ragnarok (Old Norse)
- The Saga of Mary (Old Norse)

Awards and Grants

- NSF **Graduate Student Research Fellowship** (2022)
- **Student Travel Grant** to attend DELFOL workshop at ACL, presented by Naver Labs (2019)
- **Mellon Grant** for senior thesis work, presented by Benjamin Franklin College at Yale University (2019)
- **Grace Hopper Prize** for computer science finalist (2017)
- Yale College **freshman rap battle champion** (2016)
- **Rising Scientist Award** presented by the Child Mind Institute (2015)
- **National Merit Scholarship** letter of commendation (2013)
- **Study of American History Award** presented by the Society of Mayflower Descendants (2013)
- National Latin Exam ***cum honore maximo egregio*** (2010)

Selected Coursework

Undergraduate Coursework at Yale

AI/NLP	Selected Topics in Neural Networks
	Advanced Natural Language Processing
	Computational Vision and Biological Perception
	Neural Networks and Language
	Natural Language Processing
	Deep Learning Theory and Applications
	Computing Meanings
CS Theory	Computational Complexity Theory
	Computability and Logic
	Systems Programming and Computer Organization
	Design and Analysis of Algorithms
	Data Structures and Programming Techniques
Linguistics	Formal Foundations of Linguistic Theory
	Syntax I
	Semantics I
	Phonology I
	Hybrid Grammars
	Indo-European Linguistics
	Old English
	Advanced Old English Seminar: Beowulf
Math	Introduction to Analysis
	Vector Calculus and Linear Algebra I
	Vector Calculus and Linear Algebra II

Languages

- *Coding* Python, Java, C, Haskell, PyTorch, AllenNLP, *inter alias*
- *Modern* English (Native), Icelandic (Intermediate)
- *Ancient* Latin, Old Norse, Old English