William Merrill

https://lambdaviking.com/

willm[æt]nyu.edu

Last updated September 22, 2023

Research Interests

Broad Applications of the following to deep learning, NLP, and linguistics:

- Formal languages and automata
- Computational complexity and computability
- Formal semantics

Specific Two key problems I have worked on are:

- Expressive power and inductive biases of neural nets for implementing algorithms, processing linguistic structure, and reasoning
- The theory of learning linguistic meaning from text corpora

Experience

Allen Institute for AI	2023	**Research Intern**, AllenNLP
Google Research	2022	**Student Researcher**, Speech & Lang. Algorithms
New York University	2021–	**Ph.D.**, Center for Data Science
Allen Institute for AI	2019–21	**PYI** (predoc. researcher), AllenNLP
Yale University	2015–19	**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
Google	2018	**Software Engineering Intern**
		"Exceeds expectations" rating; return offer
Boston College	2017	**Research Intern**, Language Learning Lab
New York University	2013–15	**Research Intern**, Morphology Lab

Academic Group Affiliations

CapLab, NYU	*Tal Linzen*	2021–
ML^2, NYU	*Sam Bowman, Kyunghyun Cho, He He, João Sedoc*	2021–
AllenNLP, AI2	*Noah A. Smith, Yoav Goldberg, Roy Schwartz*	2019–21
CLAY, Yale	*Robert Frank, Dana Angluin*	2016–19
L^3, Boston College	*Joshua Hartshorne, Sven Dietz*	2017
MorphLab, NYU	*Alec Marantz, Phoebe Gaston*	2013–15

Archival Publications

[1] **W. Merrill**. Formal Languages and the NLP Black Box. *Developments in Language Theory*. Ed. by F. Drewes and M. Volkov. Cham: Springer Nature Switzerland, 2023.

[2] **W. Merrill** and A. Sabharwal. The Parallelism Tradeoff: Limitations of Log-Precision Transformers. *TACL* (June 2023).

[3] **W. Merrill**, N. Tsilivis, and A. Shukla. A Tale of Two Circuits: Grokking as Competition of Sparse and Dense Subnetworks. *ICLR Workshop on Mathematical and Empirical Understanding of Foundation Models*. 2023.

[4] Z. Wu, **W. Merrill**, H. Peng, I. Beltagy, and N. A. Smith. Transparency Helps Reveal When Language Models Learn Meaning. *TACL* (2023).

[5] **W. Merrill**, A. Sabharwal, and N. A. Smith. Saturated Transformers are Constant-Depth Threshold Circuits. *TACL* (Aug. 2022).

[6] **W. Merrill**, A. Warstadt, and T. Linzen. Entailment Semantics Can Be Extracted from an Ideal Language Model. *CoNLL*. Abu Dhabi, United Arab Emirates (Hybrid): Association for Computational Linguistics, Dec. 2022.

[7] S. Subramanian, **W. Merrill**, T. Darrell, M. Gardner, S. Singh, and A. Rohrbach. ReCLIP: A Strong Zero-Shot Baseline for Referring Expression Comprehension. *ACL*. Dublin, Ireland: Association for Computational Linguistics, May 2022.

[8] M. Gardner, **W. Merrill**, J. Dodge, M. Peters, A. Ross, S. Singh, and N. A. Smith. Competency Problems: On Finding and Removing Artifacts in Language Data. *EMNLP*. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021.

[9] **W. Merrill**, Y. Goldberg, R. Schwartz, and N. A. Smith. Provable Limitations of Acquiring Meaning from Ungrounded Form: What Will Future Language Models Understand? *TACL* (Sept. 2021).

[10] **W. Merrill**, V. Ramanujan, Y. Goldberg, R. Schwartz, and N. A. Smith. Effects of Parameter Norm Growth During Transformer Training: Inductive Bias from Gradient Descent. *EMNLP*. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021.

[11] **W. Merrill**, G. Weiss, Y. Goldberg, R. Schwartz, N. A. Smith, and E. Yahav. A Formal Hierarchy of RNN Architectures. *ACL*. Online: Association for Computational Linguistics, July 2020.

[12] L. L. Wang, K. Lo, Y. Chandrasekhar, R. Reas, J. Yang, D. Burdick, D. Eide, K. Funk, Y. Katsis, R. M. Kinney, Y. Li, Z. Liu, **W. Merrill**, P. Mooney, D. A. Murdick, D. Rishi, J. Sheehan, Z. Shen, B. Stilson, A. D. Wade, K. Wang, N. X. R. Wang, C. Wilhelm, B. Xie, D. M. Raymond, D. S. Weld, O. Etzioni, and S. Kohlmeier. CORD-19: The COVID-19 Open Research Dataset. *ACL Workshop on NLP for COVID-19*. Online: Association for Computational Linguistics, July 2020.

[13] **W. Merrill**. Sequential Neural Networks as Automata. *ACL Workshop on Deep Learning and Formal Languages*. Florence: Association for Computational Linguistics, Aug. 2019.

[14] **W. Merrill**, L. Khazan, N. Amsel, Y. Hao, S. Mendelsohn, and R. Frank. Finding Hierarchical Structure in Neural Stacks Using Unsupervised Parsing. *ACL Workshop BlackboxNLP*. Florence, Italy: Association for Computational Linguistics, Aug. 2019.

[15] **W. Merrill**, G. Stark, and R. Frank. Detecting Syntactic Change Using a Neural Part-of-Speech Tagger. *ACL Workshop on Computational Approaches to Historical Language Change*. Florence, Italy: Association for Computational Linguistics, Aug. 2019.

[16] Y. Hao, **W. Merrill**, D. Angluin, R. Frank, N. Amsel, A. Benz, and S. Mendelsohn. Context-Free transductions with neural stacks. English. *EMNLP Workshop BlackboxNLP*. Brussels, Belgium: Association for Computational Linguistics, Nov. 2018.

[17] J. Kasai, R. Frank, P. Xu, **W. Merrill**, and O. Rambow. End-to-End Graph-Based TAG Parsing with Neural Networks. *NAACL*. 2018.

Non-Archival Publications

[18] **W. Merrill** and A. Sabharwal. *A Logic for Expressing Log-Precision Transformers*. Jan. 2023.

[19] M. Zhang, O. Press, **W. Merrill**, A. Liu, and N. A. Smith. *How Language Model Hallucinations Can Snowball*. June 2023.

[20] **W. Merrill** and N. Tsilivis. *Extracting Finite Automata from RNNs Using State Merging*. Jan. 2022.

[21] **W. Merrill**. *On the Linguistic Capacity of Real-Time Counter Automata*. Sept. 2020.

[22] **W. Merrill**. A semantics of subordinate clauses using delayed evaluation. *Toronto Undergraduate Linguistics Conference* (2018).

Invited Talks

- **CNRS**, Linguae Seminar, 2023
  Entailment Semantics Can Be Extracted from an Ideal Language Model
- **ICGI**, Conference Invited Speaker, 2023
  Formal Languages and Neural Models for Learning on Sequences
- **Developments in Language Theory**, Conference Invited Speaker, 2023
  Formal Languages and the NLP Black Box
- **NYC Philosophy of Language Workshop**, Invited Speaker, 2023
  Entailment Semantics Can Be Extracted from an Ideal Language Model
- **NYU**, Depth Qualifying Exam, 2023
  Transformer Reasoning Through the Lens of Circuit Complexity
- **NYU**, Guest Speaker (Comp. Ling. & Cognitive Science), 2023
  Entailment Semantics Can Be Extracted From an Ideal Language Model

- **EMNLP**, TACL Track, 2022
  Saturated Transformers are Constant-Depth Threshold Circuits

- **CoNLL**, 2022
  Entailment Semantics Can Be Extracted From an Ideal Language model

- **Microsoft Research**, New York, 2022
  The Parallelism Tradeoff: Insights on the Power and Limitations of Transformers Using Circuit Complexity

- **Umeå University**, Foundations of Language Processing, 2022
  Entailment Semantics Can Be Extracted from an Ideal Language Model

- **ArthurAI**, Journal Club, 2022
  Entailment Semantics Can Be Extracted from an Ideal Language Model

- **FLaNN Discord**, Weekly Seminar, 2022
  Saturated Transformers are Constant-Depth Threshold Circuits

- **Umeå University**, Foundations of Language Processing, 2022
  Saturated Transformers are Constant-Depth Threshold Circuits

- **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**, Speech and Language Algorithms, 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 Research 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 Evaluation

Poster Presentations

- **Philosophy of Deep Learning Workshop**, NYU, 2023
  Entailment Semantics Can Be Extracted from an Ideal Language Model
- **EMNLP**, ML Track, 2021
  Provable Limitations of Acquiring Meaning from Ungrounded Form: What Will Future Language Models Understand?
- **ACL**, Deep Learning and Formal Languages, 2019
  Sequential Neural Networks as Automata
- **ACL**, Blackbox NLP, 2019
  Finding Hierarchical Structure in Neural Stacks Using Unsupervised Parsing

Teaching Experience

University Level

- **Lead TA** for *Natural Language Processing*, Tal Linzen (NYU, Fall 2022)
- **TA** for introductory NLP (NYC AI School, Spring 2022)
- **TA** for *Artificial Intelligence*, Dragomir Radev (Yale, Spring 2019)
- **TA** for *Natural Language Processing*, Dragomir Radev (Yale, Fall 2018)
- **TA** for *Artificial Intelligence*, Dragomir Radev (Yale, Spring 2017)

High School Level and Below

- Instructor for CodeHaven (Yale, 2016-2018)
- Designed and taught *Viking Runes* (Yale Splash, Spring 2017)
- Taught *The Politics of Skyrim* (Yale Splash, Spring 2016)
- Designed and taught *DECLASSIFIED: The History of Codebreaking* (Yale Splash, Fall 2016)

SERVICE

Reviewing

NeurIPS	July 2023	*Conference*	1 review
JMLR	June 2023	*Journal*	1 review
ACL SRW	May 2023	*Workshop*	2 reviews
ICGI	April 2023	*Conference*	2 reviews
ACL	Feb 2023	*Conference*	1 review
Proc. of Royal Society A	Jan 2023	*Journal*	1 review
ARR	Nov 2022	*Conference*	1 review
Inverse Scaling Prize	Sept 2022	*Competition*	7 reviews
TheoretiCS	July 2022	*Journal*	1 review
ARR	April 2022	*Conference*	1 review
ARR	Jan 2022	*Conference*	2 review
ARR	Dec 2021	*Conference*	3 reviews
ARR	Nov 2021	*Conference*	1 review
CL	2021	*Journal*	1 review
ACL	2021	*Conference*	6 reviews
EACL	2021	*Conference*	4 reviews
EMNLP	2020	*Conference*	2 reviews
Neural Networks	2020	*Journal*	1 review

Review Excerpt (Proceedings of the Royal Society A):

> *We thank the Referee for their very thorough and constructive report on our work. It is an honor to receive such a report! We have also thanked them in the acknowledgements of our work.*

Session Chairing

ICGI	July 2023
DLT	June 2023

Formal Languages and Neural Nets (FLaNN) Community

- Scheduled and hosted weekly talk series (Fall 2022)
- Moderator of Discord server

Other

NYC AI School	2022	Volunteer instructor
AllenNLP Hackathon	2021	Technical support
AllenNLP Tutorial	2020	Chapter author
Yale Tangut Language Workshop	2018	Workshop facilitator
Yale NACLO	2017	Student volunteer
Yale Kitan Language Workshop	2016	Workshop facilitator
CodeHaven	2016–18	Student volunteer
Splash at Yale	2016–17	Student instructor

Selected Public Software

- **AllenNLP**: Open-source NLP framework (contributor)
- **The Book of Thoth**: Puzzle game with compositional spell casting in Middle Egyptian hieroglyphs
- **DraftNet**: Dota 2 drafting using neural networks
- **Voynich2Vec**: Word embedding analysis of the Voynich manuscript
- **StackNN**: Differentiable stacks, queues, and dequeues in PyTorch

Blog Posts

Research Content

- *A Formal Hierarchy of RNN Architectures* (2020)
- *Theory of Saturated Neural Networks* (2019)
- *The State of Interpretability in NLP* (2019, outdated!)
- *Word2vec Analysis of the Voynich Manuscript* (2018)
- *Review: Learning to Transduce with Unbounded Memory* (2018)
- *Capsule Networks for NLP* (2018)

Translations

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

Awards and Grants

- First annual **Angluin Invited Tutorial Speaker** (ICGI 2023)
- 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

Theoretical Computer Science and Formal Languages

- Inference and Representation (NYU, 2022)
- Foundations of Machine Learning (NYU, 2022)
- Computational Complexity Theory (Yale, 2018)
- Computability and Logic (Yale, 2017)
- Design and Analysis of Algorithms (Yale, 2017)
- Computing Meanings (Yale, 2016)
- Introduction to Computer Science (Yale, 2015)
- Formal Foundations of Linguistic Theory (Yale, 2015)

Deep Learning and Natural Language Processing

- Seminar: Scaling Laws, the Bitter Lesson, and AI Research (NYU, 2021)
- Ph.D. Introduction to Data Science (NYU, 2021)
- Seminar: Selected Topics in Neural Networks (Yale, 2019)
- Seminar: Advanced Natural Language Processing (Yale, 2018)
- Computational Vision and Biological Perception (Yale, 2018)
- Neural Networks and Language (Yale, 2018)
- Deep Learning Theory and Applications (Yale, 2018)
- Natural Language Processing (Yale, 2017)

Other Linguistics

- Hybrid Grammars: Language Contact and Change (Yale, 2019)
- Phonology I (Yale, 2018)
- The Voynich Manuscript (Yale, 2018)
- Indo-European Linguistics (Yale, 2018)
- Syntax I (Yale, 2017)
- Seminar: Beowulf and the Northern Heroic Tradition (Yale, 2017)
- Medieval Latin Paleography (Yale, 2016)
- Semantics I (Yale, 2016)
- Old English (Yale, 2015)

Other Computer Science

- Big Data (NYU, 2022)
- Systems Programming Techniques and Computer Organization (Yale, 2017)
- Data Structures and Programming Techniques (Yale, 2016)

Continuous Math

- Introduction to Analysis (Yale, 2017)
- MATH 231: Vector Calculus and Linear Algebra II (Yale, 2016)
- MATH 230: Vector Calculus and Linear Algebra I (Yale, 2015)

Reading Groups

- Nonlinear Dynamical Systems (AI2, 2021)
- Deep Learning Theory (AI2, 2020)

Languages

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