

Jacob N. Collard
1195 Valley View Road · Harpers Ferry, WV 25425 · USA
+1 740 816 7889
jacob@thorsonlinguistics.com
jacobcollard.com · github.com/thorsonlinguistics

Education

Cornell University, Ithaca, NY

PhD
2015–2020 Linguistics; Concentration: Computational Linguistics, Syntax, Semantics
MA awarded Spring 2018
PhD awarded Summer 2020
Dissertation: *A Model of Unsupervised Formal Learning for Natural Language*

Swarthmore College, Swarthmore, PA

BA
2008–2015 Linguistics, High Honors
Minor: Computer Science
Thesis: *The Linguistic Effects of Language Regulation*

Pitzer College in Botswana

Spring 2014 Study Abroad

Awards and Honors

2021 Department of Commerce Bronze Award
2019, 2015 Sage Fellowship
2016–2018 Graduate School Conference Travel Grant, Cornell University
2016–2019 Linguistics Department Conference Travel Grant, Cornell University
2015 Phi Beta Kappa

Papers

2022 J. Collard, V. de Paiva, B. Fong, E. Subrahmanian. *Extracting Mathematical Concepts from Text*. October, 2022. Workshop on Noisy User-Generated Text.

2020 J. Collard, T. N. Bhat, E. Subrahmanian, I. Monarch, J. Tash, R. Sriram, J. Elliot. *A Web Resource for Exploring the COVID-19 Dataset Using Root- and Rule-Based Phrases*. October, 2020. Journal of the Indian Institute of Science. DOI: <https://doi.org/10.1007/s41745-020-00193-2>.

2020 J. Collard. *Unsupervised Formal Grammar Induction with Confidence*. January, 2020. In *Proceedings of the Third Annual Meeting of the Society for Computation in Linguistics*. Association for Computational Linguistics. <https://scholarworks.umass.edu/scil/vol3/iss1/19>.

2018 J. Collard. *Finite State Reasoning for Presupposition Satisfaction*. August, 2018. In *Proceedings of the First International Workshop on Language, Cognition and Computational Models*. Association for Computational Linguistics. 53–62. <https://www.aclweb.org/anthology/W18-4106>.

2018 J. Collard, T. N. Bhat, E. Subrahmanian, R. Sriram, J. Elliott, U. Kattner, C. Campbell, I. Monarch. Winter, 2018. *Generating domain terminologies using root- and rule-based terms*. Journal of the Washington Academy of Sciences.

2018 T. N. Bhat, J. Collard, E. Subrahmanian, I. Monarch, J. Elliott, P. R. Strickland, B. McMahon. 2018. *R&R – a de novo method to create search terms for IUCr documents*. Acta Crystallographica. A74.

Invited Talks

2020 *Learning for Linguistics: Advancing Linguistic Theory with Computational Models of Learning*
Miami University
Oxford, Ohio
January 27, 2020

Presentations

2018 *A naturalistic Inference Learning Algorithm*
Annual Meeting of the Linguistic Society of America
Salt Lake City, USA
January 4–7, 2018
Poster, Refereed by abstract

2015 *Inference Learning for Categorical Grammars*
Logical Aspects of Computational Linguistics
Nancy, France
December 7, 2016
Talk, refereed by abstract

2015 *Natural Language Processing with SpaCy*
Cornell University
October 6, 2015
Cornell Linguistics Circle workshop

2015 *Natural Language Processing for Sociolinguistic Analysis*
International Conference on Language Documentation and Conservation
University of Hawai'i at Mānoa
February 25, 2015
Talk, refereed by abstract

2013 *Visualizing Endangered Language Contexts*
Alberta Conference on Linguistics
University of Calgary
November 2, 2013
Talk, refereed by abstract

Teaching

Spring 2021 **Adjunct Professor**, Semantics and Pragmatics
Taught graduate students in applied linguistics and computational linguistics core concepts, theories, and methods in semantics and pragmatics.

Spring 2018 **Teaching Assistant**, Introduction to Cognitive Science
Graded essays and examinations and assisted students both one-on-one and in groups.

Fall 2017 **Teaching Assistant**, Introduction to Linguistics
Ran tri-weekly discussion sessions on introductory linguistic concepts; developed and presented a guest lecture on computational linguistics.

Spring 2017 **Teaching Assisant**, Computational Linguistics
Developed and taught weekly programming sections and designed programming assignments using OCaml, FOMA, and other linguistic technologies.

Fall 2016

Instructor, Language, Thought, and Reality: The Death of Language
Developed and taught a first-year writing seminar on issues in language documentation and conservation with a practical focus on academic writing.

Related Employment

2020–Present

Post-Doctoral Researcher

National Institute of Standards and Technology
Developing and evaluating linguistically-motivated methods for representing natural language text for search and discovery. Developing tools for representing and studying linguistic knowledge in scientific text. Studying computational methods for inferring linguistic representations from text.

2021

Associate Professor

Montclair State University
Taught Semantics and Pragmatics and guided students.

2020

Consultant

Prairies to Woodlands Indigenous Language Revitalization Circle
Developing and providing technical consultation for an online dictionary and mobile app for the Michif language.

2018–2019

Research Assistant, PI: Sarah Murray

Developed a web site and deployment system for a corpus of the Cheyenne Language, including tools for rapid development of new tools for other less-studied language corpora and dictionaries.

2014–2019

Computational Linguist, National Institute of Standards and Technology

Developed methods for measuring the performance and usability of automatic terminology generation and using linguistic models to create automatically-generated terminologies from scientific publications.

2013

Undergraduate Research Assistant, Swarthmore College Endangered Languages Laboratory

Audio segmentation, alignment, and transcription, and creation of talking dictionaries for Kapingamarangi and Mokilese

2012–2013

Linguistic Data Analyst, LINGUIST List

Web design, map digitization for the [LLMap](#) project, data collection for the [Endangered Languages Project](#).

Patents

2020

Knowledge Management System and Process for Managing Knowledge, Patent Number 10,872,122

With John Elliott, Talapady N. Bhat, Ursula R. Kattner, Carelyn E. Campbell, Ram D. Sriram, Eswaran Subrahmanian, and Ira Monarch

Service and Affiliations

2020–Present

Artificial Intelligence Community of Interest
Member

2021–Present

Technical Language Processing Community of Interest
Member

2018–2019

Cornell Linguistics Circle
President

2018–2019

Cornell Interdisciplinary Semantics Research Organization
President

2018–2019

Cornell Computational Linguistics Reading Group
President & Co-Founder

2017–2018 North East Linguistic Society 49th Annual Meeting
Organizing Committee & Web Administrator

2017–Present Linguistic Society of America
Member

2016–2018 Semantics and Linguistic Theory (SALT)
Senior Editor

2015–2020 Cornell Computational Linguistics Lab
Member

2015–2020 Cornell Natural Language Processing Group
Member

2015–2018 Cornell Linguistics Circle
Web Administrator

2016–2017 Semantics and Linguistic Theory (SALT)
Assistant Editor

Other Skills

Computer Skills Programming: Python, Rust, Haskell, OCaml, C++, C
Machine Learning Frameworks: PyTorch
Web Development: HTML, CSS, SQL, JavaScript, Neo4j, Django
Typesetting: L^AT_EX

Languages Native Speaker: English
Advanced: German, Spanish
Intermediate: Mandarin, Polish
Beginner: Hindi, Norwegian, Setswana, Korean