

Jacob Collard

Computational Epistemologist

🏠 jacobcollard.com ✉ jacob@jacobcollard.com 💻 [jacobcollard](https://github.com/jacobcollard)

The key thread underlying my work is *computational epistemology*; I study how knowledge is formed, communicated, and represented formally. This begins with linguistics, the interpretive lens through which most of human knowledge is expressed, but extends into a number of related fields including logic, semiotics, and artificial intelligence. Though I work at the foundational level, studying the basic processes that make knowledge work, I strive to consider the practical, real-world impacts of all of my work, which has led to the development of computational tools that have been applied to industrial standards development, scientific research practices, and engineering.

EDUCATION

PhD in Linguistics

Cornell University · Ithaca, New York

2015–2020

- Concentration in Computational Linguistics and Semantics
- MA awarded 2018
- **Dissertation:** *A model of unsupervised formal learning for natural language*

BA in Linguistics with High Honors

Swarthmore College · Swarthmore, Pennsylvania

2011–2015

- Minor in Computer Science

RELATED EMPLOYMENT

Computer Scientist

National Institute of Standards and Technology · Gaithersburg, Maryland

2020–Present

- Developed foundational, consensus-driven, model-based standards in IoT, AI, computer security, and cloud computing.
- Designed, developed, and maintained software to process formal linguistic knowledge in standards and identify contradictions and inconsistencies.
- Collaborated with industry leaders in IoT, AI, computer security, and cloud computing to develop meaningful, theory-driven taxonomies and conceptual models.
- Published theoretical papers on mathematical knowledge graphs, technical language processing, and computational epistemology to support industry.
- Wrote successful applications for internal and external grants.
- Mentored undergraduate research projects and led team development efforts that resulted in published papers and software.
- Reviewed and edited academic papers as part of internal and external review processes.

Contractor

National Institute of Standards and Technology · Remote

2013–2020

- Developed software tools to extract domain terminologies and analyze linguistic structures in materials science, mathematics, and computer security.
- Collaborated on research projects applying computational linguistics to computer security, systems engineering, and mathematics.
- Supported ontology development and knowledge graph construction across NIST research groups and data repositories.

Adjunct Professor

Montclair State University · Remote

January, 2021–June, 2021

- Taught semantics and pragmatics, including topics such as lambda calculus, speech act theory, vector models, neural networks, inference and implicature, and cognitive science.
- Guided students through individual research projects in topics including signed languages, language and art, and discourse analysis.
- Guided students through career prospects in linguistics and computer science.

Consultant

Prairies to Woodlands Indigenous Language Revitalization Circle · Remote

February, 2022–September, 2022

- Developed full-stack web infrastructure for interactive query and navigation of multi-modal linguistic data across more than 15,000 structured lexical entries and audio archives.
- Delivered technical training on digital data management and linguistic tools to community stakeholders.
- Designed automated ETL and normalization pipelines to convert unstructured data into standardized relational dictionary schemas.

Research Assistant

Cornell University · Ithaca, New York

2018–2019

- Designed corpus storage infrastructure and feature-based search indexing for Cheyenne language resources, enabling structured queries across morphological annotations, audio, and example text.

Linguistic Data Analyst

LINGUIST List · Ypsilanti, Michigan

2012–2013

- Digitized, structured, and validated geospatial vector datasets for the LL-MAP GIS project.
- Normalized metadata schemas and database records for the Endangered Languages Project repository.
- Built web applications and database tooling to streamline internal data workflows, mailing lists, and review pipelines.

SELECTED PUBLICATIONS

Collard, Jacob, Spencer Breiner, and Eswaran Subrahmanian. 2026. "Workshop Report on Technologies and Use Cases for Smart Standards." NIST Interagency/Internal Report (NISTIR), National Institute of Standards, Technology, Gaithersburg, MD, August. <https://doi.org/https://doi.org/10.6028/NIST.IR.8622>.

Breiner, Spencer, Jacob Collard, Frederic de Vault, Eric Simmon, Eswaran Subrahmanian, and Priam Varin. 2025. *Smart Development for Smart Standards*. 66 (3). <https://oiml.org/en/publications/oiml-bulletin/2025-03/20250304>.

Collard, Jacob, Eswaran Subrahmanian, Spencer Breiner, and Ram D. Sriram. 2025. "Standards as Discourse." *International Design Conference*.

Collard, Jacob, Valeria de Paiva, and Eswaran Subrahmanian. 2024. "Mathematical Entities: Corpora and Benchmarks." In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, edited by Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. ELRA, ICCL. <https://aclanthology.org/2024.lrec-main.966>.

Collard, Jacob, Valeria de Paiva, Brendan Fong, and Eswaran Subrahmanian. 2022. "Extracting Mathematical Concepts from Text." In "Proceedings of the Eighth Workshop on Noisy User-Generated Text (W-NUT 2022)." Special issue, *Proceedings of the Eighth Workshop on Noisy User-Generated Text (W-NUT 2022)* (Gyeongju, South Korea), October, 15–23. <https://aclanthology.org/2022.wnut-1.2>.

Monarch, Ira, Jacob Collard, Sangjin Shin, Eswaran Subrahmanian, Talapady N. Bhat, and Ram Sriram. 2022. "Making Semantic Structures Explicit: Developing and Evaluating Tools and Techniques to Support Understanding of Large Cybersecurity Corpora." NIST Interagency/Internal Report (NISTIR), National Institute of Standards, Technology, Gaithersburg, MD, December. <https://doi.org/https://doi.org/10.6028/NIST.IR.8414>.

Collard, Jacob. 2020. "Unsupervised Formal Grammar Induction with Confidence." In *Proceedings of the Third Annual Meeting of the Society for Computation in Linguistics (Scil 2020)*, edited by Allyson Ettinger, Gaja Jarosz, and Joe Pater, *Proceedings of the Third Annual Meeting of the Society for Computation in Linguistics (Scil 2020)*. Association for Computational Linguistics. <https://aclanthology.org/2020.scil-1.46/>.

Collard, Jacob, Talapady Bhat, Eswaran Subrahmanian, et al. 2020. "A Web Resource for Exploring the CORD-19 Dataset Using Root- and Rule-Based Phrases." *Journal of the Indian Institute of Science* 100 (4): 725–31. <https://doi.org/10.1007/s41745-020-00193-2>.

Elliot, John, Talapady Bhat, Ursula Kattner, et al. 2020. "Knowledge Management System and Process for Managing Knowledge."

Bhat, Talapady, John Elliot, Ursula Kattner, et al. 2018. *Generating Domain Terminologies Using Root- and Rule-Based Terms*. no. 104. https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=919688.

Collard, Jacob. 2018. "Finite State Reasoning for Presupposition Satisfaction." In *Proceedings of the First International Workshop on Language Cognition and Computational Models*, edited by Manjira Sinha and Tirthankar Dasgupta, *Proceedings of the First International Workshop on Language Cognition and Computational Models*. Association for Computational Linguistics. <https://aclanthology.org/W18-4106>.

INVITED TALKS

Linguistic Modeling for Smart Standards

Technologies and Use Cases for Smart Standards · Gaithersburg, MD

March 18, 2026

Design as Discourse _____ Design Theory Workshop · Paris, France
February 3, 2025

Standards as Discourse _____ Technical Language Processing Community of Interest · Gaithersburg, MD
September 24, 2024

Linguistic Theory of Design at the Interfaces _____ Design Theory Workshop · Paris, France
January 30, 2024

Advancing Linguistic Theory with Computational Models of Learning _____ Miami University · Miami, OH
January 27, 2020

REFEREED PRESENTATIONS

The Semantics and Pragmatics of Standards _____ Information Technology Laboratory Science Day · Gaithersburg, MD
December 19, 2024

Extracting Mathematical Concepts from Text _____ LREC-COLING 2024 · Turin, Italy
May 20-25, 2024

Mathematical Concept Extraction _____ Linking Lexicographic and Language Learning Resources (4LR) · Vienna, Austria
September 15, 2023

A Naturalistic Inference Learning Algorithm _____ Linguistic Society of America · Salt Lake City, UT
January 4-7, 2018

Inference Learning for Categorical Grammars _____ Logical Aspects of Computational Linguistics · Nancy, France
December 7, 2016

TEACHING

Semantics and Pragmatics _____ Adjunct Professor · Montclair State University
2021

Introduction to Cognitive Science _____ Graduate Teaching Assistant · Cornell University
2018

Introduction to Linguistics _____ Graduate Teaching Assistant · Cornell University
2017

Computational Linguistics _____ Graduate Teaching Assistant · Cornell University
2017

Language, Thought, and Reality _____ Graduate Instructor · Cornell University
2016

SERVICE AND AFFILIATIONS

Standards Committees

National Information Standards Organization _____ Expert
2026-Present

ISO/IEC JTC 1/SC 42 Artificial Intelligence 2025-2026	Expert
ISO/IEC JTC 1/SC 41 Internet of Things 2025-2026	Expert
ISO/IEC JTC 1/SC 38 Cloud Computing 2025-2026	Expert
Academic and Professional Societies Digital Twin Consortium 2024-Present	Member
Linguistic Society of America 2017-Present	Member
Technical Language Processing Community of Interest 2021-Present	Member
Artificial Intelligence Community of Interest 2020-Present	Member
Cornell Linguistics Circle 2018-2019	President
Journal Editing and Conference Organization Workshop on Technologies and Use Cases for Smart Standards 2026	Organizing Committee
49th Annual Meeting of the North East Linguistic Society 2018	Organizing Committee
Semantics and Linguistic Theory (SALT) 2016-2018	Senior Editor
AWARDS AND HONORS	
PEAR Accolades Nominee 2024	NIST · Gaithersburg, MD
Building the Future Award 2023	NIST · Gaithersburg, MD
Bronze Award 2021	Department of Commerce · Gaithersburg, MD
Sage Fellowship 2019	Cornell University · Ithaca, NY
Sage Fellowship 2015	Cornell University · Ithaca, NY

SKILLS

Programming	Python · Rust · Haskell · OCaml · C++ · C · Typescript · JavaScript · C# · HTML · CSS · Django · Node · SQL · Postgres · Cypher · Neo4j
Ontologies and Taxonomies	RDF · OWL · SPARQL · Neo4j · SKOS
Artificial Intelligence and Machine Learning	PyTorch · HuggingFace · RAG · Vector Databases
Editing/Publishing	L ^A T _E X · Typst · Vim/Neovim · MS Office · git
Languages	English · German