Carolyn Jane Anderson

106 Central Street
Wellesley College
Wellesley, MA 02481

Web: http://canders1.github.io
Email: carolyn.anderson@wellesley.edu

Education

Ph.D. in Linguistics. University of Massachusetts, Amherst   February 2021
Dissertation: Shifting the Perspectival Landscape
Advisors: Brian Dillon and Rajesh Bhatt

M.A. in Linguistics. University of Massachusetts, Amherst   May 2019

B.A. in Linguistics and minor in Computer Science. Swarthmore College   June 2014
High Honors in Linguistics and Natural Language Models

Positions

Wellesley College, Computer Science Department   2021–present
Assistant Professor

BBN Technologies, Analytics and Machine Intelligence Group   2018–2019
Research Intern

Fulbright Canada, University of Victoria and McGill University   2014–2015
Fulbright Research Fellow, supervised by Leslie Saxon and Jessica Coon

Grants

NSF Future of Work Medium: AI-Assisted Programming: Equipping Social and Natural Scientists for the
Future of Research.   2023-2027
Collaborative award with Arjun Guha (Northeastern), Molly Feldman (Oberlin), and Erin Teich (Wellesley).

Journal Articles

1. Raymond Li, Loubna Ben Allal, Yangtian Zi, Niklas Muennighoff, Denis Kocetkov, Chenghao Mou, Marc Marone, Christopher Akiki, Jia Li, Jenny Chim, Qian Liu, Evgenii Zheltonozhskii, Terry Yue Zhuo, Thomas Wang, Olivier Dehaene, Mishig Davaadorj, Joel Lamy-Poirier, João Monteiro, Oleh Shliazhko, Nicolas Gontier, Nicholas Meade, Armel Randy, Ming-Ho Yee, Logesh Kumar Umapathi, Jian Zhu, Benjamin Lipkin, Muhtasham Oblokulov, Zhiruo Wang, Rudra Murthy, Jason Stillerman, Siva Sankalp Patel, Dmitry Abulkhanov, Marco Zocca, Manan Dey, Zhihan Zhang, Nour Fahmy, Urvashi Bhattacharyya, Suriya Gunasekar, Wenhao Yu, Swayam Singh, Sasha Luccioni, Paulo Villegas, Maxim Kunakov, Fedor Zhdanov, Manuel Romero, Tony Lee, Nadav Timor, Jennifer Ding, Claire Schlesinger, Hailey Schoelkopf, Jan Ebert, Tri Dao, Mayank Mishra, Alex Gu, Jennifer Robinson, Carolyn Jane Anderson, Brendan Dolan-Gavitt, Danish Contractor, Siva Reddy, Daniel Fried, Dzmitry Bahdanau, Yacine Jernite, Carlos Muñoz Ferrandis, Sean Hughes, Thomas Wolf, Arjun Guha, Leandro von Werra, Harm de Vries. 2023. StarCoder: May the Source Be With You! To appear in *Transactions on Machine Learning Research*.

2. Federico Cassano, John Gouwar, Daniel Nguyen, Sydney Nguyen, Luna Phipps-Costin, Donald Pinckney, Ming-Ho Yee, Yangtian Zi, Carolyn Jane Anderson, Molly Q Feldman, Arjun Guha, Michael Greenberg, Abhinav Jangda. 2023. MultiPL-E: A Scalable and Extensible Approach to Benchmarking NL2Code for 18 Programming Languages. *IEEE Transactions on Software Engineering* 49(7).

3. Anderson, Carolyn Jane and Brian Dillon. 2023. Grammatical perspective-taking in comprehension and production. *Open Mind* 7, 31-78.

4. Silva-Robles, Fe, Felipe H. Lopez, John Duff, and Carolyn Jane Anderson. 2022. Eliciting Associated Motion Constructions in Two Zapotec Languages. *Semantic Fieldwork Methods*.

5. Anderson, Carolyn Jane and Brook Danielle Lillehaugen. 2016. Negation in Colonial Valley Zapotec. *Transactions of the Philological Society* 114(3).

Peer-Reviewed Conference Proceedings Papers

1. Hannah Babe, Sydney Nguyen, Yangtian Zi, Arjun Guha, Molly Q Feldman, and Carolyn Jane Anderson. 2024. StudentEval: a Benchmark of Student-Written Prompts for Large Language Models of Code. To be appear in Association for Computational Linguistics (ACL) Findings.

2. Molly Feldman and Carolyn Jane Anderson. 2024. Non-Expert Programmers in the Generative AI Future. To appear in the *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work* (CHIWORK).

3. Sydney Nguyen, Hannah Babe, Yangtian Zi, Arjun Guha, Carolyn Jane Anderson, and Molly Q Feldman. 2024. How Beginning Programmers and Code LLMs (Mis)read Each Other. To appear in the *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*.

4. Zhao, Jingmiao and Carolyn Jane Anderson. Solving and Generating NPR Sunday Puzzles with Large Language Models. Short paper to appear in *Proceedings of the International Conference on Computational Creativity (ICCC)*.

5. Kolisko, Skylar and Carolyn Jane Anderson. 2023. Exploring Social Biases of Large Language Models in a College Artificial Intelligence Course. *Proceedings of the Thirteenth Symposium on Educational Advances in Artificial Intelligence (EAAI-23)*, co-located at AAAI 2023.

6. Phipps-Costin, Luna, Carolyn Jane Anderson, Michael Greenberg, and Arjun Guha. 2021. Solver-based Gradual Type Migration. *Proceedings of the ACM SIGPLAN Conference on Object Oriented Programming, Systems, Languages and Applications* (OOPSLA) 2021. **Selected for oral presentation. Distinguished Artifact Award.**

7. Anderson, Carolyn Jane. 2021. Tell Me Everything You Know: A Conversation Update System for the Rational Speech Acts Framework. *Proceedings of the Society for Computation in Linguistics* (SCiL) 4.

8. Anderson, Carolyn Jane and Brian Dillon. 2019. Guess who's coming (and who's going): bringing perspective to the Rational Speech Acts framework. *Proceedings of the Society for Computation in Linguistics* (SCiL) 2.

9. Anderson, Carolyn Jane, Nate Foster, Arjun Guha, Jean-Baptiste Jeannin, Dexter Kozen, Cole Schlesinger, and David Walker. 2014. NetKAT: Semantic foundations for networks. *Proceedings of the ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages* (POPL) 2014. **Selected for oral presentation.**

Peer-Reviewed Workshop Proceedings Papers

1. Ben Allal, Loubna, Raymond Li, Denis Kocetkov, Chenghao Mou, Christopher Akiki, Carlos Munoz Ferrandis, Niklas Muennighoff, Mayank Mishra, Alex Gu, Manan Dey, Logesh Kumar Umapathi, **Carolyn Jane Anderson**, Yangtian Zi, Joel Lamy Poirier, Hailey Schoelkopf, Sergey Troshin, Dmitry Abulkhanov, Manuel Romero, Terry Yue Zhuo, Francesco De Toni, Bernardo García del Río, Qian Liu, Shamik Bose, Urvashi Bhattacharyya, Michael Lappert, Ian Yu, Paulo Villegas, Jia Li, David Lansky, Huu Nguyen, Danish Contractor, Luis Villa, Daniel Fried, Dzmitry Bahdanau, Yacine Jernite, Sean Hughes, Arjun Guha, Harm de Vries, Leandro von Werra. 2022. SantaCoder: Don't Reach For the Stars! Deep Learning for Code Workshop (DL4C), 2023. **Best Paper Award**.

2. Masis, Tessa and Carolyn Jane Anderson. ProSPer: Probing Human and Neural Network Language Model Understanding of Spatial Perspective. 2021. *Proceedings of the BlackboxNLP workshop at the Conference on Empirical Methods in Natural Language Processing* 2021.

Linguistics Proceedings Papers

1. Carolyn Jane Anderson and Arjun Guha. 2024. Protagonist-Mediated Perspective. *Proceedings of Sinn und Bedeutung* (SuB) 28.

2. Bolaños Lewen, Carina and Carolyn Jane Anderson. 2022. (Some) parentheses are focus-sensitive operators. *Proceedings of Sinn und Bedeutung* (SuB) 26.

3. Anderson, Carolyn Jane. 2021. Coming in, or going out? Measuring the effect of discourse factors on perspective prominence. *Proceedings of Experiments in Linguistic Meanings* (ELM) 1.

4. Anderson, Carolyn Jane. 2019. *Tomorrow* is not always a day away. *Proceedings of Sinn und Bedeutung* (SuB) 23. **Selected for oral presentation**.

Invited Contributions

1. Guerzhoy, Michael, Marion Neumann, Pat Virtue, **Carolyn Jane Anderson**, Yaman K Singla, Alexi Orchard, Rajiv Ratn Shah. 2023. EAAI-23 Blue Sky Ideas in Artificial Intelligence Education from the AAAI/ACM SIGAI New and Future AI Educator Program. *AI Matters* 9(2).

Conference Presentations

1. Carolyn Jane Anderson and Yoolim Kim. What parenthesized modifiers (may) mean: Experimental evidence from English and Korean. Poster presented at the 37th Annual Conference on Human Sentence Processing (HSP).

2. Carolyn Jane Anderson and Yoolim Kim. Parenthesized Modifiers in English and Korean: What They (May) Mean. Poster to be presented at Experiments in Linguistic Meaning (ELM) 3.

3. Hannah Babe, Sydney Nguyen, Yangtian Zi, Arjun Guha, Molly Q Feldman, and Carolyn Jane Anderson. StudentEval: a Benchmark of Student-Written Prompts for Large Language Models of Code. Presented at the First International Workshop on Large Language Models for Code (LLM4Code), collocated with ICSE 2024. **Best Presentation Award**.

4. Federico Cassano, Luisa Li, Akul Sethi, Noah Shinn, Abby Brennan-Jones, Jacob Ginesin, Edward Berman, George Chakhnashvili, Anton Lozhkov, **Carolyn Jane Anderson**, Arjun Guha. 2024. Can

It Edit? Evaluating the Ability of Large Language Models to Follow Code Editing Instructions. Presented at the First International Workshop on Large Language Models for Code (LLM4Code), collocated with ICSE 2024.

5. Funing Yang and Carolyn Jane Anderson. Using Austen's Characters to Evaluate Computational Pipelines for Literature. Talk given at New Directions in Analyzing Text as Data 2023.

6. Cheryl Wang and Carolyn Jane Anderson. Exploring shifts in women-directed stereotypes across Wikipedia time periods. Poster presented at New Directions in Analyzing Text as Data 2023.

7. Kiara Liu, Miranda Zhu, and Carolyn Jane Anderson. XiaoshuoNLP: An Computational Pipeline for Processing Chinese Literary Texts. Poster presented at New Directions in Analyzing Text as Data 2023.

8. Carolyn Jane Anderson and Yoolim Kim. Cross-linguistic differences in processing parentheticals between English and Korean. Talk given at Comparative Punctuation Worldwide, 2023.

9. Carolyn Jane Anderson and Arjun Guha. Protagonist-Mediated Perspective. Poster presented at Sinn und Bedeutung 28.

10. Nguyen, Sydney and Carolyn Jane Anderson. Do All Minority Languages Look the Same to ChatGPT? Linguistic (Mis)information in a Large Language Model. Poster presented at the Society for Computation in Linguistics (SCiL) 2023.

11. Carolyn Jane Anderson. Making AI Education More Interdisciplinary Through Computational Creativity. Talk given at the AAAI/ACM New and Future Educators Blue Sky session at EAAI, 2023.

12. Carolyn Jane Anderson. Protagonist-Mediated Perspective. Talk given at the Narration in Context workshop at the Deutsche Gesellschaft für Sprachwissenschaft (DGfS), 2022.

13. Anderson, Carolyn Jane. Diagnosing the semantics of perspectival expressions. Poster presented at the Annual Meeting of the Linguistic Society of America (LSA), 2021.

14. Anderson, Carolyn Jane and Tessa Masis. Can neural network language models understand spatial perspective? Paper presented at the Bridging AI and Cognitive Science (BAICS) workshop, at the International Conference on Learning Representations (ICLR), 2020.

15. Anderson, Carolyn Jane and Brian Dillon. Taking other perspectives into account: an RSA model of perspectival reasoning. Talk given at Rational Approaches in Language Science (RAiLS), 2019.

16. Anderson, Carolyn Jane. Explaining the progressive motion verb puzzle in Zapotec. Talk given at the Texas Linguistics Society, 2019.

17. Anderson, Carolyn Jane. *Tomorrow* isn't always a day away. Poster presented at the 31st annual CUNY Human Sentence Processing Conference, 2018.

18. Anderson, Carolyn Jane. The San Lucas Quiaviní Zapotec andative and venitive. Talk presented at the Society for the Study of Indigenous Languages of the Americas (SSILA) annual meeting, 2018. **Honorable Mention, Best Student Paper Presentation**

19. Anderson, Carolyn Jane. The andative and venitive construction in San Lucas Quiaviní Zapotec. Talk presented at Multi-Verb Constructions: Semantic, Syntactic and Typological Perspectives (MVC), 2017.

20. Anderson, Carolyn Jane and Brook Danielle Lillehaugen. The morphosyntax of negation in Colonial Valley Zapotec. Talk presented at the Society for the Study of Indigenous Languages of the Americas (SSILA) annual meeting, 2015.

21. Anderson, Carolyn Jane and Brook Danielle Lillehaugen. La morfosintaxis de la negation en el zapoteco del Valle colonial. Talk presented at Coloquio sobre Lenguas Otomangues y Vecinas (COLOV) IV: Mario Molina Cruz, 2014.

22. Anderson, Carolyn Jane. Language ideology and human rights doctrine in Morocco. Talk presented at New Ways of Analyzing Variation (NWAV) 42, 2013.

Manuscripts

1. Federico Cassano, Luisa Li, Akul Sethi, Noah Shinn, Abby Brennan-Jones, Jacob Ginesin, Edward Berman, George Chakhnashvili, Anton Lozhkov, **Carolyn Jane Anderson**, Arjun Guha. 2024. Can It Edit? Evaluating the Ability of Large Language Models to Follow Code Editing Instructions.

2. Federico Cassano, John Gouwar, Francesca Lucchetti, Anders Freeman, Claire Schlesinger, Carolyn Jane Anderson, Molly Q Feldman, Michael Greenberg, Abhinav Jangda, and Arjun Guha. 2023. Knowledge Transfer from High-Resource to Low-Resource Programming Languages for Code LLMs.

3. Funing Yang and Carolyn Jane Anderson. 2023. Using Austen's Characters to Evaluate Computational Pipelines for Literature.

4. Anderson, Carolyn Jane. 2021. Untangling classes of context-sensitivity: a closer look at the semantics of American English *tomorrow*.

5. Anderson, Carolyn Jane. 2021. Shifting the Perspectival Landscape: Methods for Encoding, Identifying, and Selecting Perspectives. Dissertation, University of Massachusetts, Amherst.

6. Anderson, Carolyn Jane. 2017. The andative and venitive construction in San Lucas Quiaviní Zapotec. Ms., University of Massachusetts, Amherst.

7. Anderson, Carolyn Jane. 2014. I talk it and I feel it: Language attitudes of Moroccan university students. Honors thesis, Swarthmore College.

Invited Talks

"Which Programmers Will Generative AI Empower? Lessons from Charlie the Coding Cow."

- Wellesley College Reunion Faculty Lecture Series. 2024.

"Modeling Perspectives in Narratives."

- Stories that Win Symposium hosted by the Incubator for Transdisciplinary Futures at Washington University. 2024.

"Demystifying Large Language Models."

- Wellesley College Alumnae Association Faculty Speaker Series. 2024.

"Perspectives, Puzzles, and Programming: Linguistics Meets Computer Science."

- Swarthmore College Linguistics Colloquium. 2024.

"Evaluating Large Language Models of Code."

- Brookhaven National Laboratory AI/ML Seminar Series. 2023.

"Large Language Models v. Humans."

- University of Connecticut School of Law. 2023.

"Evaluating multi-language code generation."

- MathWorks. 2023.

"A Bayesian approach to modeling grammatical perspective-taking."

- GdR LIFT research group, ILFC (les interactions entre linguistiques formelles et computationnelles) Seminar. 2023.
- Brandeis University, Department of Linguistics and Computational Linguistics. 2021.

"Guess Who's Coming to Dinner: a Bayesian approach to modeling perspective."

- Boston University, Department of Linguistics. 2020.
- Cornell Tech, Department of Computer Science. 2020.
- Colgate University, Department of Computer Science. 2019.
- Denison University, Department of Data Analytics. 2019.
- James Madison University, Department of Computer Science. 2019.
- Mount Holyoke College, Department of Computer Science. 2019.
- Occidental College, Department of Computer Science. 2019.
- University of Massachusetts, Amherst, Department of Computer Science. 2019.
- Wellesley College, Department of Computer Science. 2019.

Other Presentations

1. Anderson, Carolyn Jane. The Future of Programming. Talk given in the Wellesley College Science Center Faculty Seminar Series, 2024.
2. Anderson, Carolyn Jane. What (Some) Parentheses Mean. Poster presented at the UMass Linguistics department's 50th anniversary celebration, 2023.
3. Anderson, Carolyn Jane. Reasoning about perspectives. Abstract selected for presentation at the Science Speed Geeking event at the XPrag Summer School, 2019.
4. Anderson, Carolyn Jane. Use your words (better): two projects on making better use of text for ASR. Intern presentation, BBN Technologies, 2019.
5. Anderson, Carolyn Jane. Neural networks for ASR: focusing on the text. Intern presentation, BBN Technologies, 2018.

Teaching

1. Instructor, Wellesley College, Computer Science Department   Fall 2022 and Fall 2023
   CS333: Natural Language Processing
2. Instructor, Wellesley College, Computer Science Department Spring 2022, Fall 2022, Fall 2023, Spring 2024
   CS232: Artificial Intelligence
3. Instructor, Wellesley College, Computer Science Department   Fall 2021, Spring 2022, Spring 2023, Spring 2024
   CS111: Introduction to Computer Science

4. Instructor, Wellesley College, Computer Science Department   Fall 2021
   CS251: Principles of Programming Languages

5. Instructor, University of Massachusetts Amherst, Linguistics Department   Fall 2019
   First Year Seminar: Technology for Language Revitalization

6. Teaching Assistant, University of Massachusetts Amherst, Linguistics Department Fall 2019
   Language Processing and the Brain (Instructor: Brian Dillon)

7. Instructor, Mt. Holyoke College, Computer Science Department   Fall 2018
   Programming Languages

8. Teaching Assistant, University of Massachusetts Amherst, Linguistics Department Spring 2017
   Introduction to Linguistic Theory (Instructor: Seth Cable)

9. Teaching Assistant, University of Massachusetts Amherst, Linguistics Department Fall 2017
   Cognitive Modeling (Instructor: Gaja Jarosz)

Advising

Wellesley College Research Projects

Zixuan Wu   2024
Pattern recognition in text-vision models

Abby Brennan-Jones   2023-2024
How do beginning programmers use code generation tools?

Yuneydy Paredes   2023-2024
How do beginning programmers use code generation tools?

Hangyeol Park (Co-advised with Yoolim Kim)   2023-2024
Project: Parentheticals in Korean and English

Jin Ryu (Co-advised with Yoolim Kim)   2023-2024
Project: Parentheticals in Korean and English

Sydney Nguyen   2022-2023
How do beginning programmers use code generation tools?

Rachelle Hu   2022-2023
The geometric space of code embeddings

Kiara Liu and Miranda Zhu   2022-2023
Building an NLP pipeline for Chinese literature

Cheryl Wang   2022-2023
Quantifying word embedding bias from historical events

Darcy Kim, Aizah Rao, Alexandra Chin   2021-2022
Quantifying word embedding bias from historical events

Kiara Liu   2021-2022
Inspector BERT: murder mysteries as a case study on language model narrative understanding

Angel Ding   2021-2022
Machine learning tools for identifying and classifying marginalia in medieval manuscripts

Rachelle Hu   2021-2022
How do beginning programmers use code generation tools?

Wellesley College Honors Theses

Anders Freeman   2023-2024
The Quest for Quality Data

Michelle Zhao   2022-2023
Solving and Generating NPR Sunday Puzzles with Large Language Models

Skylar Kolisko   2022
Exploring Name Bias in NLP Models

Funing Yang   2021-2022
An Extraction and Representation Pipeline for Literary Characters

University of Massachusetts, Amherst

Tessa Masis   2019-2020
Project: Experimental and computational modeling of perspectival motion verbs

Alicia LeClair   2017
Project: Perspectival motion verbs in Western Tlacolula Valley Zapotec

Other

Carina Bolaños Lewen   2020-2021
Project: Semantics and pragmatics of perspectival motion verbs and parentheticals

Awards

AAAI/ACM SIGAI New and Future AI Educator   2023

Computing Research Association Outstanding Undergraduate Researchers Award Finalist   2014

The Linguistics Prize in Applications of Theory, Swarthmore College   2014

Phi Beta Kappa Inductee   2014

Philip M. Hicks Prize for Best Critical Essay   2013

Fellowships

BBN Technologies Research Fellowship   2019 and 2020

Selkirk Linguistics Outreach Fund Award   2016

University of Massachusetts Amherst Fellowship   2015–2016

Fulbright Canada Research Scholarship   2014–2015

Programming Languages Mentoring Workshop Scholarship   2013 and 2014

Marianne Durand Frey '57 Scholar   2011–2014

National Merit Scholar   2010

Research Assistantships

University of Massachusetts Amherst, Linguistics Department   2017
Research Assistant
Primary investigator: Seth Cable
Topic: Semantics of Endangered Languages in North America

University of Massachusetts Amherst, Linguistics Department   2016
Research Assistant
Primary investigator: Gaja Jarosz
Topic: Computational phonology

Haverford College, Ticha Digital Humanities Project   2014
Research Intern
Advisor: Brook Lillehaugen
Topic: Web-based preservation and presentation of Colonial Valley Zapotec texts

Cornell University, Computer Science Department   2013
Research Intern, supervised by Nate Foster
Topic: NetKAT: a domain-specific language for networking

Service

Internal Service

Library Awards Committee committee member   2024

Extramural Graduate Fellowship committee member   2023-2024

Ruhlman committee member   2023, 2024

CS representative to the Summer Research Program selection committee   2022-2024

Tenure-track advisory committee member   2022-2023

External Service

Area Chair, EMNLP, Linguistic Theories, Cognitive Modeling, and Psycholinguistics track   2023

Editorial Board, *Semantics & Pragmatics*   2023-present

Reviewer, *Communications of the ACM*   2023, 2024

Reviewer, *Society for Computation in Linguistics*   2022, 2023, 2024

Reviewer, *Society for Human Sentence Processing* (HSPS) Annual Conference   2022, 2023, 2024

Reviewer, BlackboxNLP workshop at EMNLP   2022, 2023, 2024

Reviewer, *Open Mind*   2024

Reviewer, NeurIPS Datasets and Benchmarks track   2023

Reviewer, *Association for the Advancement of Artificial Intelligence* (AAAI)   2022

Reviewer, ACL Rolling Review   2021

Reviewer, *Psychological Review*   2021

Reviewer, *Language*   2020