

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).

Publications

1. 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*.
2. 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, Urvasi 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*.
3. 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)*.
 4. 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).
 5. 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**.
 6. Anderson, Carolyn Jane and Brian Dillon. 2023. Grammatical perspective-taking in comprehension and production. *Open Mind* 7, 31-78.
 7. 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).
 8. 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.
 9. Silva-Robles, Fe, Felipe H. Lopez, John Duff, and Carolyn Jane Anderson. 2022. Eliciting Associated Motion Constructions in Two Zapotec Languages. *Semantic Fieldwork Methods*.
 10. Bolaños Lewen, Carina and Carolyn Jane Anderson. 2022. (Some) parentheses are focus-sensitive operators. *Proceedings of Sinn und Bedeutung (SuB)* 26.
 11. 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.
 12. 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**.
 13. 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.
 14. 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.
 15. Anderson, Carolyn Jane. 2019. Tomorrow is not always a day away. *Proceedings of Sinn und Bedeutung (SuB)* 23. **Selected for oral presentation**.

16. 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.
17. Anderson, Carolyn Jane and Brook Danielle Lillehaugen. 2016. Negation in Colonial Valley Zapotec. *Transactions of the Philological Society* 114(3).
18. 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.**

Conference Presentations

1. 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. To be presented at the First International Workshop on Large Language Models for Code (LLM4Code), collocated with ICSE 2024.
2. 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.
3. 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.
4. 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.
5. Carolyn Jane Anderson and Yoolim Kim. Cross-linguistic differences in processing parentheticals between English and Korean. Talk given at Comparative Punctuation Worldwide, 2023.
6. Carolyn Jane Anderson and Arjun Guha. Protagonist-Mediated Perspective. Poster presented at Sinn und Bedeutung 28.
7. Nguyen, Sydney and Carolyn Jane Anderson. Do All Minority Languages Look the Same to Chat-GPT? Linguistic (Mis)information in a Large Language Model. Poster presented at the Society for Computation in Linguistics (SCiL) 2023.
8. 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.
9. Carolyn Jane Anderson. Protagonist-Mediated Perspective. Talk given at the Narration in Context workshop at the Deutsche Gesellschaft für Sprachwissenschaft (DGfS), 2022.
10. Anderson, Carolyn Jane. Diagnosing the semantics of perspectival expressions. Poster presented at the Annual Meeting of the Linguistic Society of America (LSA), 2021.
11. 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.
12. 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.
13. Anderson, Carolyn Jane. Explaining the progressive motion verb puzzle in Zapotec. Talk given at the Texas Linguistics Society, 2019.

14. Anderson, Carolyn Jane. *Tomorrow* isn't always a day away. Poster presented at the 31st annual CUNY Human Sentence Processing Conference, 2018.
15. 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**
16. 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.
17. 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.
18. 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.
19. Anderson, Carolyn Jane. Language ideology and human rights doctrine in Morocco. Talk presented at New Ways of Analyzing Variation (NWAV) 42, 2013.

Manuscripts

1. Funing Yang and Carolyn Jane Anderson. 2023. Using Austen's Characters to Evaluate Computational Pipelines for Literature.
2. Anderson, Carolyn Jane. 2021. Untangling classes of context-sensitivity: a closer look at the semantics of American English *tomorrow*.
3. Anderson, Carolyn Jane. 2021. Shifting the Perspectival Landscape: Methods for Encoding, Identifying, and Selecting Perspectives. Dissertation, University of Massachusetts, Amherst.
4. Anderson, Carolyn Jane. 2017. The andative and venitive construction in San Lucas Quiaviní Zapotec. Ms., University of Massachusetts, Amherst.
5. Anderson, Carolyn Jane. 2014. I talk it and I feel it: Language attitudes of Moroccan university students. Honors thesis, Swarthmore College.

Invited Talks

"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. What (Some) Parentheses Mean. Poster presented at the UMass Linguistics department’s 50th anniversary celebration, 2023.
2. Anderson, Carolyn Jane. Reasoning about perspectives. Abstract selected for presentation at the Science Speed Geeking event at the XPrag Summer School, 2019.
3. Anderson, Carolyn Jane. Use your words (better): two projects on making better use of text for ASR. Intern presentation, BBN Technologies, 2019.
4. 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
Programming Languages | Fall 2018 |
| 8. Teaching Assistant, University of Massachusetts Amherst, Linguistics Department
Introduction to Linguistic Theory (Instructor: Seth Cable) | Spring 2017 |
| 9. Teaching Assistant, University of Massachusetts Amherst, Linguistics Department
Cognitive Modeling (Instructor: Gaja Jarosz) | Fall 2017 |

Advising

Wellesley College Research Projects

Abby Brennan-Jones How do beginning programmers use code generation tools?	2023-2024
Yunedy Paredes How do beginning programmers use code generation tools?	2023-2024
Hangyeol Park (Co-advised with Yoolim Kim) Project: Parentheticals in Korean and English	2023-2024
Jin Ryu (Co-advised with Yoolim Kim) Project: Parentheticals in Korean and English	2023-2024
Sydney Nguyen How do beginning programmers use code generation tools?	2022-2023
Rachelle Hu The geometric space of code embeddings	2022-2023
Kiara Liu and Miranda Zhu Building an NLP pipeline for Chinese literature	2022-2023
Cheryl Wang Quantifying word embedding bias from historical events	2022-2023
Darcy Kim, Aizah Rao, Alexandra Chin Quantifying word embedding bias from historical events	2021-2022
Kiara Liu Inspector BERT: murder mysteries as a case study on language model narrative understanding	2021-2022
Angel Ding Machine learning tools for identifying and classifying marginalia in medieval manuscripts	2021-2022
Rachelle Hu How do beginning programmers use code generation tools?	2021-2022

Wellesley College Honors Theses

Anders Freeman Improving Large Language Models of Code for R	2023-2024
Michelle Zhao Solving and Generating NPR Sunday Puzzles with Large Language Models	2022-2023

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 Research Assistant Primary investigator: Gaja Jarosz Topic: Computational phonology	2016
Haverford College, Ticha Digital Humanities Project Research Intern Advisor: Brook Lillehaugen Topic: Web-based preservation and presentation of Colonial Valley Zapotec texts	2014
Cornell University, Computer Science Department Research Intern, supervised by Nate Foster Topic: NetKAT: a domain-specific language for networking	2013

Service

Extramural Graduate Fellowship committee member	2023-2024
Ruhlman committee member	2023, 2024
Tenure-track advisory committee member	2022-2023
CS representative to the Summer Research Program selection committee	2022
Coding Club supervisor	2021-2022
Volunteer, Skype A Scientist	2020
Organizer, Neural Networks in Linguistics Reading Group	2018-2020

Reviewing

Editorial Board, <i>Semantics & Pragmatics</i>	2023
Area Chair, EMNLP, Linguistic Theories, Cognitive Modeling, and Psycholinguistics track	2023
Reviewer, NeurIPS Datasets and Benchmarks track	2023
Reviewer, <i>Communications of the ACM</i>	2023
Reviewer, <i>Society for Computation in Linguistics</i>	2022, 2023
Reviewer, <i>Society for Human Sentence Processing</i> (HSPS) Annual Conference	2022, 2023
Reviewer, BlackboxNLP workshop at EMNLP	2022, 2023
Reviewer, <i>Association for the Advancement of Artificial Intelligence</i> (AAAI)	2022
Reviewer, ACL Rolling Review	2021
Reviewer, <i>Psychological Review</i>	2021
Reviewer, <i>Language</i>	2020