

C.M. DOWNEY

Assistant Professor

Department of Linguistics · Goergen Institute for Data Science & AI
University of Rochester

c.m.downey@rochester.edu

cmdowney88.github.io · ur2nlp.github.io · @cmdowney

EDUCATION

University of Washington

2018–2024

Ph.D. Computational Linguistics

Advisors: Gina-Anne Levow and Shane Steinert-Threlkeld

M.S. Computational Linguistics earned upon candidacy, June 2022

Tulane University

2014–2018

B.S. Linguistics, Russian

Advisors: Charles Mignot and Judith Maxwell

Thesis: Competing Motivations Driving Semi-Optimal Focus Construction Types

EMPLOYMENT

University of Rochester

2024–present

Assistant Professor - Linguistics and Data Science

University of Washington

2018–2024

Teaching/Research Assistant

Microsoft

Summer 2022

Research Intern, MSR Cryptography and Privacy

Apple

Summer 2020

AI/ML Intern, Siri Web Answers

HONORS & AWARDS

Empire AI Beta Allocation - Rapid Adaptation of ASR Models in Data-Scarce Scenarios

2025

University of Rochester; 5,000 service units

Presidential Dissertation Fellowship

2024

UW Graduate School

Excellence in Linguistic Research Fellowship

2022

UW Department of Linguistics

Foreign Language and Area Studies Fellowship - Inuktitut Language

2019–2020

U.S. Department of Education, UW Jackson School of International Studies

Newberry Research Library Summer Institute

2019

The Newberry Library, Chicago; Fellow; Indigenous Language Revitalization

Inducted into Phi Beta Kappa Honor Society	2018
Tulane University	
Outstanding Linguistics Thesis	2018
Tulane University	
Presidential Scholarship	2014
Tulane University	

PUBLICATIONS

Peer-Reviewed

2026. [Linguistically Informed Evaluation of Multilingual ASR for African Languages](#). Fei-Yueh Chen, Lateef Adeleke, **C.M. Downey**. *To appear: Proceedings of the 7th Workshop on African Natural Language Processing (AfricaNLP)*.

2024. [Targeted Multilingual Adaptation for Low-resource Language Families](#). **C.M. Downey**, Terra Blevins, Dhwani Serai, Dwija Parikh, Shane Steinert-Threlkeld. *Findings of the Association for Computational Linguistics: EMNLP 2024*.

2023. [Embedding Structure Matters: Comparing Methods to Adapt Multilingual Vocabularies to New Languages](#). **C.M. Downey**, Terra Blevins, Nora Goldfine, Shane Steinert-Threlkeld. *Proceedings of the 3rd Workshop on Multilingual Representation Learning*.

Best Paper Award

2023. [Learning to Translate by Learning to Communicate](#). **C.M. Downey**, Leo Z. Liu, Xuhui Zhou, Shane Steinert-Threlkeld. *Proceedings of the 3rd Workshop on Multilingual Representation Learning*.

2022. [A Masked Segmental Language Model for Natural Language Segmentation](#). **C.M. Downey**, Fei Xia, Gina-Anne Levow, Shane Steinert-Threlkeld. *Proceedings of the 19th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*.

2022. [Multilingual unsupervised sequence segmentation transfers to extremely low-resource languages](#). **C.M. Downey**, Shannon Drizin, Levon Haroutunian, Shivin Thukral. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*.

2022. [Emergent Communication Fine-tuning \(EC-FT\) for Pretrained Language Models](#). Shane Steinert-Threlkeld, Leo Z. Liu, Xuhui Zhou, **C.M. Downey**. *Proceedings of the 5th Annual Workshop on Emergent Communication*.

Runner-up Best Paper Award

Dissertation

2024. [Adapting Pre-Trained Models and Leveraging Targeted Multilinguality](#). **C.M. Downey**. *PhD Dissertation, University of Washington*.

Pre-prints

2022. [Planting and Mitigating Memorized Content in Predictive-Text Language Models](#). **C.M. Downey**, Wei Dai, Huseyin A. Inan, Kim Laine, Saurabh Naik, Tomasz Religa. *Arxiv*.

TALKS

Refereed

Targeted Multilingual Adaptation for Low-resource Language Families The 2024 Conference on Empirical Methods in Natural Language Processing (poster)	2024
Embedding Structure Matters: Comparing Methods to Adapt Multilingual Vocabularies to New Languages 3rd Workshop on Multilingual Representation Learning, EMNLP (oral)	2023
Learning to Translate by Learning to Communicate 3rd Workshop on Multilingual Representation Learning, EMNLP (poster)	2023
A Masked Segmental Language Model for Natural Language Segmentation 19th SIGMORPHON Workshop, NAACL (oral)	2022
Multilingual unsupervised sequence segmentation transfers to extremely low-resource languages 60th Annual Meeting of the Association for Computational Linguistics (poster)	2022

Invited

Improving Computational Tools for Under-resourced and Endangered Languages	
University at Buffalo Department of Linguistics	2026
George Mason University Linguistics Colloquium	2025
University of Rochester Center for Integrated Research Computing Symposium	2024
University of British Columbia Department of Linguistics	2024
University of Pittsburgh Department of Linguistics	2024
Macalester College Department of Linguistics	2024
University of Utah Department of Linguistics	2024
University of Rochester Goergen Institute for Data Science and Department of Linguistics	2023
Multilingual Tokenizer Adaptation: Specializing Representations for Target Languages	2025
Speech and text LLMs revisited, Université Paris Cité	
Fostering Grassroots Language Technology and AI	2025
Goergen Institute for Data Science and AI 10th Anniversary	
Integrating NLP Models into Language Documentation Workflows	2024
Central New York Humanities Corridor Sound Structures Workshop	

TEACHING

Instructor

Spring 2026	UR DSAC 251/451: Machine Learning with Limited Data UR LING 250/450: Data Science for Linguistics
Fall 2025	UR LING 282/482: Deep Learning for Computational Linguistics
Spring 2025	UR LING 250/450: Data Science for Linguistics
Fall 2024	UR LING 282/482: Deep Learning for Computational Linguistics
Summer 2023	UW LING 473: Basics for Computational Linguistics (<i>Co-Instructor with Emily Proch Ahn</i>)
Spring 2023	UW LING 574: Deep Learning for Natural Language Processing

Teaching Assistant

Winter 2023	UW LING 572: Advanced Statistical Methods for Natural Language Processing
Spring 2022	UW LING 574: Deep Learning for Natural Language Processing
Winter 2022	UW LING 200: Introduction to Linguistic Thought
Fall 2021	UW LING 269: Swearing and Taboo Language
Spring 2021	UW LING 574: Deep Learning for Natural Language Processing
Fall 2020	UW LING 566: Introduction to Syntax for Computational Linguistics
Spring 2019	UW LING 200: Introduction to Linguistic Thought UW ASL 305: Introduction to American Deaf Culture (<i>grader</i>)
Fall 2018	UW LING 200: Introduction to Linguistic Thought UW LING 406: Introduction to Syntax (<i>grader</i>)

Guest Lectures

May 2024	“Multilingual Language Modeling” <i>LING 574: Deep Learning for NLP, University of Washington, instructed by Shane Steinert-Threlkeld</i>
May 2022	“Multilingual Language Modeling” <i>LING 574: Deep Learning for NLP, University of Washington, instructed by Shane Steinert-Threlkeld</i>
May 2021	“Multilingual Language Modeling” <i>LING 574: Deep Learning for NLP, University of Washington, instructed by Shane Steinert-Threlkeld</i>
May 2020	“Computational Linguistics and Language Revitalization” <i>LING 234: Language and Diversity, University of Washington, instructed by Lorna Rozelle</i>
Jan 2020	“Computational Linguistics and Language Revitalization” <i>ENGL 4717: NAIS Capstone Seminar, University of Colorado, instructed by Penelope Kelsey</i>

SERVICE

UR Linguistics Graduate Admissions Committee	2025
Reviewer	
National Science Foundation	2025
Panelist	
Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP)	2024, 2023
Reviewer	
Multilingual Information Access Workshop	2022
Reviewer	