

Catherine Arnett

ccarnett@ucsd.edu

<https://catherinearnett.github.io/>

Education

- PhD Candidate, Linguistics with Computational Social Science**, UC San Diego 2019-
Committee: Farrell Ackerman, Benjamin Bergen, Leon Bergen, Victor Ferreira
- MA Chinese and Linguistics (Hons)**, University of Edinburgh 2014-2018
Including one year exchange at Zhejiang University, Hangzhou, P.R. China

Teaching

- Graduate Teaching Consultant for International Instructors**, UC San Diego 2022-
Instruct incoming international graduate students in teaching skills and subject-specific English
- Instructor, LIGN 170: Psycholinguistics**, UC San Diego Summer 2023
[Course Website](#)
- Teaching Assistant**, UC San Diego 2019-2022
LIGN 101: Introduction to Language, Fall 2020-Spring 2022
LIDS 19: Independent Language Study, Winter 2020-Spring 2020
LIGN 170: Psycholinguistics, Fall 2019

Publications

Refereed Conference Proceedings

1. James A. Michaelov*, **Catherine Arnett***, Tyler A. Chang, Benjamin K. Bergen (2023). [Structural priming demonstrates abstract grammatical representations in multilingual language models](#). Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP). Singapore. *equal contribution
2. **Catherine Arnett** (2019). [Pathways of Change in Romance Motion Events: A Corpus-based Comparison](#). *Proceedings of the Thirtieth Western Conference on Linguistics (WeCOL)*. Vol 23. Fresno, CA.

Manuscripts

1. Tyler A. Chang, **Catherine Arnett**, Zhuowen Tu, Benjamin K. Bergen (under review). [When Is Multilinguality a Curse? Language Modeling for 250 High- and Low-Resource Languages](#). arXiv preprint arXiv:2311.09205.
2. **Catherine Arnett**, Tyler A. Chang, Benjamin K. Bergen (under review). [A Bit of a Problem: Measurement Disparities in Dataset Sizes Across Languages](#). arXiv preprint arXiv:2403.00686.

Conference Presentations

1. **Catherine Arnett**, Tyler A. Chang, James A. Michaelov, and Benjamin K. Bergen (2023) [Crosslingual Structural Priming and the Pre-Training Dynamics of Bilingual Language Models](#). *The 3rd Workshop on Multilingual Representation Learning co-located with EMNLP 2023*. Singapore.
2. **Catherine Arnett** & Maho Takahashi (2022). Creating a Baseline to Evaluate Correlations Between Language and Environment. *Machine Learning and the Evolution of Language*. Kanagawa/online. [[abstract](#)] [[poster pdf](#)]
3. **Catherine Arnett** & Eva Wittenberg (2020). Multiple Meanings of Doubling Up: Mandarin Verbal Reduplication. *The 26th Architectures And Mechanisms for Language Processing Conference (AMLaP)*. Potsdam/online. [[poster pdf](#)]
4. **Catherine Arnett** & Eva Wittenberg (2020). Conceptual Effects of Verbal Reduplication in Mandarin Chinese. *North American Conference on Chinese Linguistics 32*. Storrs, CT/online.
5. **Catherine Arnett** & Eva Wittenberg (2019). Conceptual Effects of Verbal Reduplication in Mandarin Chinese. California Meeting on Psycholinguistics. Santa Cruz, California. [[poster pdf](#)]
6. **Catherine Arnett** (2018). Diachronic study of the typology of motion verbs in the Romance languages. Undergraduate Linguistics Association of Britain Conference, University of Edinburgh.

Invited Talks

Writing Systems. LIGN 101: Introduction to the Study of Language Fall 2020

Media and Outreach

Interview Jan 2024

[Code and Community: Computational Social Science Program Addresses Social Questions with Data](#)

Author and Subject-Matter Expert, Sep 2022

[Cognitive Foundations, an open Textbook](#)

Update and rewrite portions of the *Language* chapter using bookdown in R.

Tools Created

Byte Premium Tool Feb 2024

<https://github.com/catherinearnett/byte-premium-tool>

Command-line tool to calculate dataset scaling needed to achieve cross-lingual training data equity.

Scholarships, Fellowships, and Grants

Yankelovich Graduate Research Funding, UC San Diego, \$3,500 2024

GPSA Travel Grant Award, UC San Diego, \$300 2023

Linguistics Department Anti-Racist Pedagogy Micro-Fellowship, UC San Diego Linguistics, \$500 2023

Summer Graduate Teaching Scholar, UC San Diego, \$1,000 2023

LaVerne Noyes Foundation Endowed Fellowship, UC San Diego, \$6,000 2022

Academic Senate Grant, UC San Diego, \$9,000 2022

Friends of the International Center Fellowship, UC San Diego, \$2000	2020
Research and Travel Grant, UC San Diego Linguistics, \$350	2019
GPSA Travel Grant Award, UC San Diego, \$500	2019
Student Opportunity Fund, University of Edinburgh, £250	2018
Eric Liddell China Saltire Scholarship, University of Edinburgh, £5,000	2016
St. Andrews Society of North Carolina Scholarship, \$37,000	2014-2018

Awards

Edinburgh Award, Peer Learning and Support	2018
--	------

Certifications

Citi Program Social and Behavioral Research Best Practices for Clinical Research	2019-2022
HSK 5 Certification	2017

Skills and Languages

Languages

- English** - Native
- Mandarin** - Fluent
- Spanish** - Fluent
- French** - Conversational
- Latin** - Structural Knowledge
- San Juan Piñas Mixtec** - Fieldwork Experience

Coding and Software

- R/RStudio** tidyverse, lme4, bookdown
- Python** pandas, matplotlib, seaborn, numpy, BeautifulSoup
- basic SQL, basic HTML
- Software** Google Suite, Microsoft Office Suite, Adobe CC, Slack, Zoom, GitHub

Interests

Natural language processing, psycholinguistics, corpus work, morphology

Academic Service

Reviewing

Undergraduate Linguistics Association of Britain	2020
--	------

Departmental Service

Linguistics Department Steward	2022-2024
Linguistics Department Graduate Student Writing Club Organizer	2022-2023
Linguistics Department Social Organizer	2021-2023
Linguistics Faculty Growth Planning Committee	2021-2022
Linguistics Department BrownBag Organizer	2020-2021
Linguistics Department Representative, Graduate and Professional Student Association	2019-2020

Mentoring

Masters Students

Adelle Engmann, Computational Social Science	2023-2024
Xinyao Yi, Computational Social Science	2022-2023

Undergraduate Honors Theses

Yuhan Fu, Cognitive Science B.S. Honors Thesis "Is There Evidence for Embodied Simulation During Language Translation?"	2024
--	------

Undergraduate Research Assistants

Stephanie Huang	Winter 2024 -
Sidney Ma	Winter 2024 -
Gwen O'Brien	Winter 2024 -
Nattida Neal	Winter 2024 -
Sarah Yu	Fall 2023 -
Livia Zhou	Fall 2023
Yoyo Wu	Fall 2023
Jiawei Lyu	Spring 2023
Jason Tran	Winter 2023 - Spring 2023
Emily Xu	Fall 2022, Spring 2023 - Fall 2023
Tiffany Wu	Fall 2022 - Winter 2023
Fiona Tang	Fall 2022 - Spring 2023
Jane Yang	Winter 2020 - Spring 2022
Zhiyi Li	Winter 2020 - Winter 2021
Xutong Zhang	Fall 2020 - Winter 2021
Annabelle Chang	Fall 2019 - Winter 2021
Jessica Luo, now Linguistics PhD Student at University of Washington	Fall 2019 - Winter 2021
Ruoqi Wei	Fall 2019 - Winter 2021
Rebecca Xu	Fall 2019 - Winter 2020

Professional Memberships

Association for Computational Linguistics

2023