

Aryaman Arora

DEPARTMENT OF COMPUTER SCIENCE
STANFORD UNIVERSITY

 aryaman.io
 aryamana@stanford.edu
 aryaman2020

EDUCATION

Stanford University, Stanford, CA 2023—
Ph.D., Computer Science
First-year rotations: Dan Jurafsky, Christopher Potts, Diyi Yang

Georgetown University, Washington, D.C. 2020—2023
B.S., Computer Science GPA: 3.796
B.A., Linguistics

EXPERIENCE

Georgetown University, Washington, D.C. Aug. 2019—Sep. 2023
Undergraduate Researcher, NERT (Dr. Nathan Schneider)

Redwood Research, Berkeley, CA Winter 2022
Research Resident

Apple, Seattle, WA Summer 2022
AI/ML Intern, Siri Natural Language

ETH Zürich, Zürich, Switzerland Summer 2021
Research Intern, Rycolab (Dr. Ryan Cotterell)

PUBLICATIONS

Vedant Palit*, Rohan Pandey*, **Aryaman Arora**, and Paul Pu Liang. 2023. **Towards vision-language mechanistic interpretability: A causal tracing tool for BLIP**. In *5th Workshop on Closing the Loop Between Vision and Language*.

Omer Goldman, Khuyagbaatar Batsuren, Salam Khalifa, **Aryaman Arora**, Garrett Nicolai, Reut Tsarfaty, and Ekaterina Vylomova. 2023. **SIGMORPHON–UniMorph 2023 shared task 0: Typologically diverse morphological inflection**. In *SIGMORPHON*.

Aryaman Arora, Adam Farris, Samopriya Basu, and Suresh Kolichala. 2023. **Jambu: A historical linguistic database for South Asian languages**. In *SIGMORPHON*.

Brett Reynolds, **Aryaman Arora**, and Nathan Schneider. 2023. **Unified syntactic annotation of English in the CGEL framework**. In *LAW*.

Khuyagbaatar Batsuren, Gábor Bella, **Aryaman Arora**, ..., and Ekaterina Vylomova. 2022. **The SIGMORPHON 2022 shared task on morpheme segmentation**. In *SIGMORPHON*.

Jordan Kodner, ..., **Aryaman Arora**, ..., and Ekaterina Vylomova. 2022. **SIGMORPHON–UniMorph 2022 shared task 0: Generalization and typologically diverse morphological inflection**. In *SIGMORPHON*.

Khuyagbaatar Batsuren*, Omer Goldman*, ..., **Aryaman Arora**, ..., and Ekaterina Vylomova. 2022. **UniMorph 4.0: Universal Morphology**. In *LREC*.

Aryaman Arora. 2022. **Universal Dependencies for Punjabi**. In *LREC*.

- Aryaman Arora**, Nitin Venkateswaran, and Nathan Schneider. 2022. *MASALA: Modelling and analysing the semantics of adpositions in linguistic annotation of Hindi*. In *LREC*.
- Aryaman Arora**, Clara Isabel Meister, and Ryan Cotterell. 2022. *Estimating the entropy of linguistic distributions*. In *ACL*.
- Aryaman Arora**, Adam Farris, Samopriya Basu, and Suresh Kolichala. 2022. *Computational historical linguistics and language diversity in South Asia*. In *ACL*.
- Adam Farris* and **Aryaman Arora***. 2021. *For the purpose of curry: A UD treebank for Ashokan Prakrit*. In *UDW, SyntaxFest*.
- Aryaman Arora**, Adam Farris, Gopalakrishnan R, and Samopriya Basu. 2021. *Bhāṣācitra: Visualising the dialect geography of South Asia*. In *LChange*.
- Aryaman Arora**, Nitin Venkateswaran, and Nathan Schneider. 2021. *SNACS annotation of case markers and adpositions in Hindi*. In *SCiL*.
- Michael Kranzlein, Emma Manning, Siyao Peng, Shira Wein, **Aryaman Arora**, and Nathan Schneider. 2020. *PASTRIE: A corpus of prepositions annotated with supersense tags in Reddit International English*. In *LAW*.
- Aryaman Arora**, Luke Gessler, and Nathan Schneider. 2020. *Supervised grapheme-to-phoneme conversion of orthographic schwas in Hindi and Punjabi*. In *ACL*.
- Aryaman Arora** and John R. McIntyre. *Quasi-passive lower and upper extremity robotic exoskeleton for strengthening human locomotion*. In *Sustainable Innovation*.

TALKS AND NON-ARCHIVAL PRESENTATIONS

- Aryaman Arora**. 2023. *Investigating induction heads in a small transformer language model*. In *MASC-SLL, Arlington, VA, USA*.
- Aryaman Arora**, Nathan Schneider, and Brett Reynolds. 2022. *A CGEL-formalism English treebank*. In *MASC-SLL, Philadelphia, PA, USA*.
- Adam Farris* and **Aryaman Arora***. 2022. *DIPI: Dependency parsing for Ashokan Prakrit historical dialectology*. In *Towards a comparative historical dialectology: evidence from morphology and syntax, DGfS, Tübingen, Germany*.
- Aryaman Arora** and Nathan Schneider. 2020. *SNACS annotation of case markers and adpositions in Hindi*. In *SIGTYP*.

PREPRINTS

- Nicholas Goldowsky-Dill, Chris MacLeod, Lucas Sato, and **Aryaman Arora**. 2023. *Localizing model behavior with path patching*. *arXiv*, arXiv:2304.05969.
- Brett Reynolds, **Aryaman Arora**, and Nathan Schneider. 2022. *CGELBank: CGEL as a framework for English syntax annotation*. *arXiv*, arXiv:2210.00394.
- Aryaman Arora** and Ahmed Etebari. 2021. *Kholosi Dictionary*. Updated 2021-04-13.

IN PREPARATION

- Aryaman Arora**, Clara Meister, Ethan Wilcox, and Ryan Cotterell. 2023. *Structured entropy estimation*. *Under review*.

TEACHING

Teaching Assistant

Intro to Comp Science: Python (COSC 010)

Fall 2022

ADVISING

Adam Farris

[B.S. CS and Math, Stanford University, 2026]

Kabilan Prasanna

[High school student, 2025]

SERVICE

Shared task co-organiser

SIGMORPHON, Morpheme Segmentation 2022

SIGMORPHON–UniMorph, Morphological Inflection 2022, 2023

Program committee, SIGTYP Workshop 2022, 2023, 2024

Tutorial, Information Theory in Linguistics: Methods and Applications (COLING) 2022

Trebank maintainer, Universal Dependencies 2021—

Contributor

UniMorph 2021—

WikiPron 2020—

Member, Association for Computational Linguistics 2020—

Problemsetter, Panini Linguistics Olympiad 2020

Administrator, English Wiktionary 2017—

AWARDS

Barry M. Goldwater Scholarship 2022

U.S. Presidential Scholars Program Candidate 2020

LANGUAGES

Native: English, Hindi–Urdu

Proficient: Mandarin Chinese, Punjabi | Python (PyTorch), C++, JavaScript, R

Fieldwork: Kholosi, Dagaare