

Aryaman Arora

DEPARTMENT OF COMPUTER SCIENCE
STANFORD UNIVERSITY

 aryaman.io

 aryamana@stanford.edu

 aryaman2020

EDUCATION

Stanford University, Stanford, CA 2023—
Ph.D., Computer Science
Advisors: Dan Jurafsky, Christopher Potts

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 (Nathan Schneider)

Redwood Research, Berkeley, CA Winter 2022
Research Resident

Apple, Seattle, WA Summer 2022
AI/ML Intern, Siri Natural Language (Robert Daland)

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

PUBLICATIONS

PREPRINTS

- [1] Zhengxuan Wu, Atticus Geiger, **Aryaman Arora**, Jing Huang, Zheng Wang, Noah D. Goodman, Christopher D. Manning, and Christopher Potts. 2024. **pyvene: A library for understanding and improving PyTorch models via interventions**. *arXiv:2403.07809*. Updated 2024-03-12.
- [2] **Aryaman Arora**, Dan Jurafsky, and Christopher Potts. 2024. **CausalGym: Benchmarking causal interpretability methods on linguistic tasks**. *arXiv:2402.12560*. Updated 2024-02-19.
- [3] Zhengxuan Wu, Atticus Geiger, Jing Huang, **Aryaman Arora**, Thomas Icard, Christopher Potts, and Noah D. Goodman. 2024. **A reply to Makelov et al. (2023)’s “interpretability illusion” arguments**. *arXiv:2401.12631*. Updated 2024-01-23.
- [4] Nicholas Goldowsky-Dill, Chris MacLeod, Lucas Sato, and **Aryaman Arora**. 2023. **Localizing model behavior with path patching**. *arXiv:2304.05969*. Updated 2023-05-16.
- [5] Brett Reynolds, **Aryaman Arora**, and Nathan Schneider. 2022. **CGELBank: CGEL as a framework for English syntax annotation**. *arXiv:2210.00394*. Updated 2022-10-01.
- [6] **Aryaman Arora** and Ahmed Etebari. 2021. *Kholosi Dictionary*. Updated 2021-04-13.

CONFERENCES

- [1] Kabilan Prasanna and **Aryaman Arora**. 2024. *IruMozhi: Automatically classifying diglossia in Tamil*. *Findings of NAACL*.
- [2] Khuyagbaatar Batsuren*, Omer Goldman*, ..., **Aryaman Arora**, ..., and Ekaterina Vylomova. 2022. *UniMorph 4.0: Universal Morphology*. In *LREC*.
- [3] **Aryaman Arora**. 2022. *Universal Dependencies for Punjabi*. In *LREC*.
- [4] **Aryaman Arora**, Nitin Venkateswaran, and Nathan Schneider. 2022. *MASALA: Modelling and analysing the semantics of adpositions in linguistic annotation of Hindi*. In *LREC*.
- [5] **Aryaman Arora**, Clara Isabel Meister, and Ryan Cotterell. 2022. *Estimating the entropy of linguistic distributions*. In *ACL*.
- [6] **Aryaman Arora**, Adam Farris, Samopriya Basu, and Suresh Kolichala. 2022. *Computational historical linguistics and language diversity in South Asia*. In *ACL*.
- [7] **Aryaman Arora**, Luke Gessler, and Nathan Schneider. 2020. *Supervised grapheme-to-phoneme conversion of orthographic schwas in Hindi and Punjabi*. In *ACL*.

WORKSHOPS

- [1] Nay San, Georgios Paraskevopoulos, **Aryaman Arora**, Xiluo He, Prabhjot Kaur, Oliver Adams, and Dan Jurafsky. 2024. *Predicting positive transfer for improved low-resource speech recognition using acoustic pseudo-tokens*. In *SIGTYP*.
- [2] 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*.
- [3] 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*.
- [4] **Aryaman Arora**, Adam Farris, Samopriya Basu, and Suresh Kolichala. 2023. *Jambu: A historical linguistic database for South Asian languages*. In *SIGMORPHON*.
- [5] Brett Reynolds, **Aryaman Arora**, and Nathan Schneider. 2023. *Unified syntactic annotation of English in the CGEL framework*. In *LAW*.
- [6] Khuyagbaatar Batsuren, Gábor Bella, **Aryaman Arora**, ..., and Ekaterina Vylomova. 2022. *The SIGMORPHON 2022 shared task on morpheme segmentation*. In *SIGMORPHON*.
- [7] Jordan Kodner, .., **Aryaman Arora**, .., and Ekaterina Vylomova. 2022. *SIGMORPHON–UniMorph 2022 shared task 0: Generalization and typologically diverse morphological inflection*. In *SIGMORPHON*.
- [8] Adam Farris* and **Aryaman Arora***. 2021. *For the purpose of curry: A UD treebank for Ashokan Prakrit*. In *UDW, SyntaxFest*.
- [9] **Aryaman Arora**, Adam Farris, Gopalakrishnan R, and Samopriya Basu. 2021. *Bhāṣācitra: Visualising the dialect geography of South Asia*. In *LChange*.
- [10] **Aryaman Arora**, Nitin Venkateswaran, and Nathan Schneider. 2021. *SNACS annotation of case markers and adpositions in Hindi*. In *SCiL*.

- [11] 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*.

PRESENTATIONS

- [1] **Aryaman Arora**. 2023. Investigating induction heads in a small transformer language model. In *MASC-SLL*, Arlington, VA, USA.
- [2] **Aryaman Arora**, Nathan Schneider, and Brett Reynolds. 2022. **A CGEL-formalism English treebank**. In *MASC-SLL*, Philadelphia, PA, USA.
- [3] 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.
- [4] **Aryaman Arora** and Nathan Schneider. 2020. **SNACS annotation of case markers and adpositions in Hindi**. In *SIGTYP*. Non-archival extended abstract.

OTHER

- [1] **Aryaman Arora** and John R. McIntyre. 2019. **Quasi-passive lower and upper extremity robotic exoskeleton for strengthening human locomotion**. In *Sustainable Innovation*.

TEACHING

Teaching Assistant

Intro to Comp Science: Python (COSC 010) Fall 2022

MENTORSHIP

Adam Farris Undergrad, Stanford University	2021—2023
Vedant Palit Undergrad, IIT Kharagpur	2023—
Kabilan Prasanna High school student	2023—
Markus Zhang Undergrad, Stanford University	2023—

SERVICE

<i>Area Chair</i> , NAACL	2024
<i>Reviewer</i> , LREC-COLING	2024
<i>Program committee</i>	
SIGTYP	2022, 2023, 2024
EURALI	2024
<i>Shared task co-organiser</i>	
SIGMORPHON, Morpheme Segmentation	2022
SIGMORPHON–UniMorph, Morphological Inflection	2022, 2023
<i>Tutorial</i> , Information Theory in Linguistics: Methods and Applications (COLING)	2022
<i>Treebank maintainer</i> , Universal Dependencies	2021—
<i>Contributor</i>	
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