Aryaman Arora

RESEARCH INTERESTS: Computational linguistics and natural language processing, especially natural language understanding (NLU) through annotation and corpus design, and computational typology/historical linguistics for South Asian languages.

EMAIL: aa2190@georgetown.com, aryaman.arora2020@gmail.com

WEBSITE: aryamanarora.github.io

TWITTER: @aryaman2020

EDUCATION

Georgetown University, Washington, D.C.   *August 2020 – Present*
B.S. in Computer Science and Linguistics (double major)   GPA: 3.91/4.00
Relevant Coursework: Empirical Methods in NLP, Computational Corpus Linguistics, Artificial Intelligence, Field Methods

RESEARCH EXPERIENCE

Member of **NERT**, **Georgetown University**, Washington, D.C.   *July 2019 – Present*
> Advised by Dr. Nathan Schneider (whose interests are in meaning representations, semantic tagging/parsing, and annotation). Published research at ACL conferences and workshops on various topics regarding Indian languages: phonology, semantic tagging, language change. Involved in collaborative work for Universal Dependencies, UniMorph, and WikiPron.

Research Intern at **Rycolab**, **ETH Zürich**, Zürich, Switzerland   *July 2021 – August 2021*
> Intern with Dr. Ryan Cotterell (who works on morphology, information theory, NLP algorithms, etc.), working on entropy estimation and information-theoretic analysis of language.

INDUSTRY EXPERIENCE

Intern at **Apple**, Seattle, Washington   *May 2022 – August 2022*
> Intern in the Siri Natural Language Understanding team.

VOLUNTEERING

Co-founder of **Washingtutors**, Washington, D.C.   *March 2020 – Present*
> Run an 501(c)(3) non-profit organisation (co-founded with Ananya Gulati) that provides free tutoring to D.C. Public School students, through online platforms, having served over 50 students.

PUBLICATIONS (REFEREED)

1. Tiago Pimentel, ... **Aryaman Arora**, ..., and Colin Wilson. 2022. A pipeline for generating plausible wug words. Under review.

2. Khuyagbaatar Batsuren*, Omer Goldman*, ..., **Aryaman Arora**, ..., and Ekaterina Vylomova. 2022. UniMorph 4.0: Universal Morphology. Under review.

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

4. **Aryaman Arora**, Nitin Venkateswaran, and Nathan Schneider. 2022. A corpus of Hindi adposition and case semantics. In *LREC*.

5. **Aryaman Arora**, Clara Isabel Meister, and Ryan Cotterell. 2022b. Estimating the entropy of linguistic distributions. In *ACL*.

6. **Aryaman Arora**, Adam Farris, Samopriya Basu, and Suresh Kolichala. 2022a. Computational historical linguistics and language diversity in South Asia. In *ACL*.

7. Adam Farris* and **Aryaman Arora***. 2021. For the purpose of curry: A UD treebank for Ashokan Prakrit. In *UDW, SyntaxFest*.

8. **Aryaman Arora**, Adam Farris, Gopalakrishnan R, and Samopriya Basu. 2021a. Bhāṣācitra: Visualising the dialect geography of South Asia. In *LChange*.

9. **Aryaman Arora**, Nitin Venkateswaran, and Nathan Schneider. 2021b. SNACS annotation of case markers and adpositions in Hindi. In *SCiL*.

10. 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*.

11. **Aryaman Arora** and Nathan Schneider. 2020. SNACS annotation of case markers and adpositions in Hindi. In *SIGTYP*.

12. **Aryaman Arora**, Luke Gessler, and Nathan Schneider. 2020. Supervised grapheme-to-phoneme conversion of orthographic schwas in Hindi and Punjabi. In *ACL*.

13. **Aryaman Arora** and John R. McIntyre. 2020. Quasi-passive lower and upper extremity robotic exoskeleton for strengthening human locomotion. In Anshu Saxena Arora, Sabine Bacouel-Jentjens, Mohamad Sepehri, and Amit Arora, editors, *Sustainable Innovation*. Palgrave Pivot.

SERVICE

Shared tasks:
- Co-organiser, UDParse-IL @ 6th Workshop on Indian Language Data: Resources and Evaluation (*2022*)
- Co-organiser, SIGMORPHON 2022 Shared Task on Morpheme Segmentation (*2022*)

Program Committee:
- Program Committee, SIGTYP Workshop (*2022*)

Other:
- Treebank Maintainer, Universal Dependencies (*2021*–)
- Contributor, UniMorph (*2021*–)
- Contributor, WikiPron (*2020*–)
- Problemsetter, Panini Linguistics Olympiad (*2020*–)
- Administrator, English Wiktionary (*2017*–)

AWARDS

- Barry M. Goldwater Scholarship (*2022*)

LANGUAGES

NATIVE: English, Hindi–Urdu
INTERMEDIATE: Punjabi
LEARNING: Mandarin, Indo-Aryan languages
FIELDWORK: Kholosi (Iran), Dagaare (Burkina Faso)

SKILLS

PROGRAMMING: Python, C++, JavaScript, R, Java, Bash
ML: PyTorch, scikit-learn
WEB: jQuery, D3.js
ANNOTATION: UD, PTB, SNACS, AMR
MARKUP: HTML + CSS, LaTeX, Markdown