

# Vasundhara Gautam

Burnaby BC      vasundharagautam.com

Canada      vasundhara131719@gmail.com

I am a computational linguist with industry and research experience in speech recognition, discourse processing, corpus linguistics and laboratory phonology. My research interests include multi-user and low-resource language and speech processing, and using insights from linguistics (especially phonetics and phonology) for NLP.

## Education

### **BSc Linguistics and Computing Science, Simon Fraser University**

*September 2015 - April 2019*

Graduated with distinction. Cumulative GPA 3.99. Upper Division GPA 4.14.

## Work Experience

### **Speech Recognition Engineer, Dialpad**

*May 2019 - Present*

- Integrated class-based and ngram language modelling in a WFST architecture, for improved speech recognition of classes of English words
- Maintained code to domain-adapt (at the word level) a generic English speech recognition model to companies on-the-fly based on metadata
- Expanded the scope of this domain adaptation to the user level, increasing the recall of targeted words without negatively impacting WER
- Maintained English lexica through manual review of entries (generated by G2P and corrected by transcribers); wrote a syllabifier to auto-flag entries that violate English syllable structure
- Scoped speech recognition for languages other than English and made a long-term plan for the company, weighing linguistic considerations, technical and data limitations, and financial constraints
- Spearheaded the creation of a policy about inclusive language at the company, both technical and otherwise

### **Software Engineering Intern, GE Digital**

*January 2018 - August 2018*

- Developed software for Predix App Engine, a platform for the Industrial Internet of Things
- added features, resolved bugs, wrote automated tests, and maintained documentation

- Led the effort to write a comprehensive technical Installation Guide for on-premise installation of our multi-node, multi-service platform, which required broad knowledge of the platform services, networking, system administration and security
- Involved in overhauling the software architecture of authentication and access control for the product using Cloud Foundry User Account and Authentication (UAA)

## Research Experience

### Research Assistant, SFU Discourse Processing Lab

*September 2018 - Present*

- Worked in a group led by Dr Maite Taboada on three projects: Gender Gap Tracker (GGT), TACT (Topic Analysis, Constructiveness and Toxicity), and Adverb-ly Adjectives
- **GGT:** This project seeks to quantify gender bias in the media by counting the proportion of people quoted in news text who are women. To do so, we developed a system that analyzes articles published online by certain English-language Canadian news outlets, using parts we built in combination with off-the-shelf NLP tools.
  - Supervised manual annotation of English corpus for evaluation of the Tracker
  - Developed the preliminary, unreleased Gender Gap Tracker for French news articles
- **TACT:** This project analyzed millions of online comments on English news articles to examine toxicity, constructiveness, and content in this genre and their interplay. Our findings yield insights for news outlets to moderate comments and advise commenters, as well as for internet citizens looking to have more meaningful interactions with others
  - Developed a Python pipeline to clean online comment data, predict toxicity and constructiveness of comments and topic distributions, and visualize the results
  - Trained an LDA topic model on online comments in Canadian news, and found that people are more constructive when they talk about certain topics, but that toxicity is an equal-opportunity offender
- **Adverb-ly Adjectives:** Aimed to create a semantic classification of the [adverb-ly adjective] construction - an adjective modified by an adverb ending in 'ly'
  - Used R and Python to computationally extract, count and classify instances of these constructions in large corpora

### Research Assistant, SFU Phonological Processing Lab

*September 2016 - April 2019*

- Worked in a group led by Dr Ashley Farris-Trimble on several projects aimed at studying phonological processing by using eye-tracking experiments in the visual world paradigm
- Curated visual and auditory stimuli, created word lists, programmed the experiments with SR Research Experiment Builder and MATLAB
- Used an eye-tracker and an audiometer to run participants in experiments
- Organized and ran linguistics-themed day camps to recruit children for studies

## Posters, Presentations and Publications

- *The Gender Gap Tracker: Using Natural Language Processing to measure gender bias in media*  
Asr, F. T., M. Mazraeh, A. Lopes, **V. Gautam**, J. Gonzales and M. Taboada (accepted to PLoS ONE, 2021).
- *How To ARPABET: introduction to phonemic transcription using ARPABET*. **Gautam, V.** (2020). Dialpad Data Science Club. October 2020. Presentation.
- *Ethics in NLP*. **Gautam, V.** (2020). Dialpad Data Science Club. July 2020. Presentation.
- *Vowels: in phonetics, phonology, acoustics and speech recognition*. **Gautam, V.** (2020). Dialpad ASR Learning. June 2020. Presentation.
- *Catchy Title Goes Here: how to make better presentations*. **Gautam, V.** (2020). Dialpad Data Science Club. May 2020. Presentation.
- *Rational, Constructive Discourse in Online Comment Sections*  
**Gautam, V.** (2019). Redeye, Vancouver Co-op Radio. November 2019. Radio interview.
- *Hey Tyee Commenters! Scholars Studied You. Here's What They Found*  
**Gautam, V.** and M. Taboada (2019). The Tyee. November 2019. Article.
- *Intermediate RegEx Practice*. **Gautam, V.** (2019). Dialpad ASR Learning. November 2019. Presentation.
- *Flagging Mispronunciations: introduction to phonology and syllabification*. **Gautam, V.** (2019). Dialpad ASR Learning. October 2019. Presentation.
- *Technical Changes and User Impact*. **Gautam, V.** (2019). Dialpad ASR Learning. August 2019. Presentation.
- *Automating Comment Moderation: Topics and Toxicity in Online Comments*  
**Gautam, V.** and M. Taboada (2019). Linguistics Poster Session, Simon Fraser University. April 2019. Poster.
- *Formant Transitions: Tense versus Lax Vowels*  
**Gautam, V.**, J. Gonzalez and B. Jeon (2019). Topics in Phonetics, Simon Fraser University. April 2019. Presentation.
- *Automating Comment Moderation: Topics and Toxicity in Online Comments*  
**Gautam, V.** and M. Taboada (2019). Undergraduate Research Symposium, Simon Fraser University. March 2019. Presentation.
- *Language-Independent Named Entity Recognition*  
**Gautam, V.** (2018). AI Poster Session, Simon Fraser University. December 2018. Poster.
- *Content moderation in social media*  
Taboada, M. and **V. Gautam** (2018). Symposium on Public Discourse and Data Science, Simon Fraser University. October 2018. Presentation.

## Scholarships and Awards

### **Outstanding Academic Achievement Award** (Spring 2019)

Awarded at convocation by the Department of Linguistics for consistent academic excellence

### **Gordon M. Shrum National Entrance Scholarship** (Fall 2015 - Spring 2019)

Full-ride scholarship conditional on continued academic excellence

**Big Data Undergraduate Student Research Award** (Spring 2019)

Awarded to work part-time with Dr Maite Taboada at the Discourse Processing Lab

**Social Sciences and Humanities Undergraduate Student Research Award** (Fall 2018)

Awarded to work part-time with Dr Maite Taboada at the Discourse Processing Lab

**Mark and Nancy Brooks Computing Science Innovation Award** (Fall 2018)

Granted semesterly to a Computing Science student who "demonstrates exceptional accomplishment, promise or innovation in computing science outside classroom work"

**Inspiring Women in Applied Sciences Award** (Fall 2018)

Granted annually to a female student with "a demonstrated history of community engagement activities of benefit to women in STEM-related fields"

**Dean's List**

Awarded in every eligible term for a GPA of 3.5 or higher

**President's List**

Awarded in 4/5 eligible terms for a GPA of 4.0 or higher

## Extra-Curricular Activities

- Birding
- Indian classical music (Carnatic and Hindustani styles)
- Indian classical dance (Bharatanatyam style)
- Photography
- Languages (native - English; advanced - Hindi, Tamil, Malayalam; conversational - French, Spanish; beginner - Italian, Mandarin, ASL)

## Service

- Co-founded ColourID, a Dialpad Employee Resource Group for employees who are Black, Indigenous and people of colour
- Invited speaker at the Queer in AI social held at COLING 2020 (December 2020)
- Queer in AI organizer for EMNLP 2020, AACL 2020 and COLING 2020 (September 2020 - Present)
- Invited speaker for CS Education Week at Stride Avenue Community School: led interactive sessions about programming with 6 classes of children from grades 1-6 (November 2020)
- EMNLP 2020 reviewer (August 2020)
- Young Women in Business WISE program workshop series: Taught a 4-hour workshop on VBA for Excel (March 2020)
- Women in Computing Science: Try/Catch Conference, Data Science Mentor (May 2019)
- Perform for Pride: A Benefit Concert for Qmunity, Master of Ceremonies (January 2019)
- Girls Learning Code, Coding Mentor (March 2016 - December 2018)
- SFU Residence Hall Association, VP communications (May 2017 - August 2017)
- Women in Computing Science: Try/Catch Conference, Photographer (May 2017)
- Kids Code Jeunesse, Coding Mentor (March 2016 - June 2016)
- SFU Residence Hall Association, House Representative (October 2015 - April 2016)