

Katherine Blake

Data Scientist, Amazon Alexa



kmlake95@gmail.com



317 496 7841



[katherineblake.github.io](https://github.com/katherineblake)

Coursework

(Graduate-level, Cornell University)
Phonetics, Phonology, Syntax,
Semantics, Computational
Linguistics, Linguistic Typology,
Methodology of Romance Language
Learning and Teaching, Historical
Linguistics

(Computer Science, Cornell
University) Introduction to Computing
with Python (CS1110, Spring 2018),
Object-Oriented Programming & Data
Structures (CS2110, Fall 2018); [Deep
Learning Specialization](#) (Coursera)

Skills

Programming languages: Python
(Proficient); Java, MATLAB, SQL
(Familiar)

Natural languages: English (Native);
French (Proficient); Italian, Spanish
(Intermediate)

Programs: \LaTeX , Praat, MS Office

Other: Phonetic transcription (IPA)

Education

2017 - 2022 Doctor of Philosophy in Linguistics
Cornell University | Ithaca, NY

Dissertation:

Phonological markedness effects on noun-adjective ordering

Advisor: Dr. Marten van Schijndel

2013 - 2016 Bachelor of the Arts in Linguistics with Honors, Highest Distinction
Indiana University | Bloomington, IN

Undergraduate Thesis: *Rhoticity in Cajun French*

Linguistics

Advisor: Dr. Kelly H. Berkson

2013 - 2016 Bachelor of the Arts in French with Highest Distinction
Indiana University | Bloomington, IN
Spring 2015 at Université d'Aix-Marseille | Aix-en-Provence, France

Work Experience

2022-present Language Data Scientist at Amazon Alexa Seattle, WA
Conduct quantitative and qualitative analyses of natural language
data and human annotation for the evaluation and continuous im-
provements of machine learning (ML) and large language models
(LLMs); Project management, data selection and design of human an-
notation tasks to create training data for LLMs.

Summer 2021 Language Data Research Intern at Amazon Alexa Remote
Conducted quantitative and qualitative analyses of large text corpora
to build datasets for machine learning models that improve routing
behavior of customer requests

2020 - 2021 Sr. Computational/Programmer Analyst at [Army Research Lab](#)
Contributed to ongoing projects in neural text-to-speech synthesis
and machine translation (evaluation metrics and NER annotation);
co-instructor of the visible speech course for [GEMS summer program](#)

2019 - 2020 Intern at [Synfonica Speech, LLC](#) Ithaca, NY
Did copy synthesis using Praat to improve text-to-speech output
of the company's rule-based synthesizer, wrote English phonotactic
rules for speech therapy software (Python), organized code

2018 - 2021 Graduate Teaching Assistant at Cornell University Ithaca, NY
Led discussion sections, evaluated student work, created lesson and
test materials for Introduction to Linguistics, Introduction to Phonet-
ics & Phonology, and First Year French courses.

Selected Presentations

- Blake, Katherine. (2022). Avoiding phonological markedness via word order-
ing in French and Italian. Talk at the 2022 Linguistics Society of America (LSA)
Annual Meeting. Washington, D.C./Online.
- Al Sagheer, Z., Blake, K., Iglehart, V., LaRocca, S., Morgan, J., Murray, J. (2020).
Shareable TTS Components. Talk at the 2020 American Association for Ma-
chine Translation (AMTA) conference. Online.

Recognition

2017, 2021 Sage Fellowship Cornell University
Tuition + stipend for first year and dissertation year of PhD

May 2017 Induction into Phi Beta Kappa Indiana University
National Honors Society

2017 Provost's Award for Undergraduate Research Indiana University
Social and Applied Science

2013 - 2017 Undergraduate of Highest Distinction Indiana University
All undergraduate career, top academic honors