

Kristen Sheets

Greater Boston, MA Region

EDUCATION

(415) 233-3032 · sheetskristen@gmail.com · Github/LinkedIn: \sheetskristen

M.S. in Computational Linguistics

expected 2020

Brandeis University; Waltham, MA

Coursework: Fundamentals of Computational Linguistics, Linguistic Annotation for Machine Learning, Statistical & Neural Methods for NLP, Computational Semantics, NLP Systems for Named Entity Recognition, Python for NLP, Java Programming, Data Structures & Algorithms

Teaching assistantships: Linguistic Typology, Syntactic Theory (*for Prof. Lotus Goldberg*)

B.A. in Linguistics

2014

University of California, Santa Cruz; Santa Cruz, CA

Coursework: Syntax V, Semantics II, Phonology II, Phonetic Analysis, Language Typology, Morphology

Teaching assistantships: Syntactic Structures (*for Prof. Maziar Toosarvandani*)

Education abroad: Institut d'études politiques; Lyon, France

SKILLS

Programming: Python (NLTK, spaCy, sklearn, pandas), Java, Haskell, Bash, regex, SQL, L^AT_EX

Natural languages: English (native), German (fluent), French (proficient)

WORK EXPERIENCE

Research Assistant

September 2018 - *present*

PI: Prof. James Pustejovsky, Brandeis University; Waltham, MA

- Researches inherent polysemy using corpora and data-driven methodologies within a Generative Lexicon framework in the Lab for Linguistics and Computation.

Linguistic Data Intern

June 2019 - August 2019

Basis Technology; Cambridge, MA

- Created process and methodologies for in-house natural language processing tool evaluation, using open source tools including SpaCy and Kuromoji, and executed on a large comparison project in collaboration with the R&D team.
- Sourced novel multilingual data sets leveraging Wikipedia, census data and other publicly available resources, through the augmentation and generation of custom data processing scripts using Python and Bash.

Solutions Engineer

December 2017 - July 2018

Ericsson Emodo (formerly Placecast); San Francisco, CA

- Designed and automated data integration processes by writing and maintaining process controllers for data ingestion, primarily in Bash.
- Served as a technical expert to internal and external stakeholders for custom technical integrations, producing a wide range of deliverables including data extraction and analysis for Product, Engineering and Business Development teams.

Technical Analyst

January 2016 - December 2017

Ericsson Emodo (formerly Placecast); San Francisco, CA

- Executed data analysis, requirement engineering, functionality testing and technical documentation tasks for internal and external stakeholders, including Fortune 100 enterprise partners.

Localization Project Coordinator

July 2014 - January 2016

Google; Mountain View, CA

- Managed end to end global localization production, including triaging functional and linguistic issues for multiple product lines including Gmail, Google Apps for Work, and Google Play across 68 languages.
- Trained new Localization Project Coordinators, managing ramp and development for 5 new team members.

SELECTED PROJECTS

Named Entity Recognition for Podcasts

Developing a custom ontology, extending from the OntoNotes NER annotation guidelines, for named entity recognition and extraction from podcast transcripts. Future work to include topic modeling using extracted named entities.

Semantic Type Classifier

Using the WordNet annotated SemCor corpus, I designed and trained a classification model to identify a word's CoreLex basic semantic type, as a proof of concept for the task identifying trends in polysemy. I'm currently expanding on this work, building a neural model using contextual word embeddings (BERT).