

# Derek C. Andersen

## Computational Linguist

✉ [derek@derekandersen.net](mailto:derek@derekandersen.net)    [dechrissen](#)    [dechrissen.github.io](#)    [derekcandersen](#)

## EXPERIENCE

### NLU Developer

#### Cerence Inc.

 2021 – Present    Remote

- Improved the accuracy of AI-based language models for use in automotive voice assistant technology (data synthesis, bug fixing, error analysis).
- Wrote Python scripts for data processing and task automation.
- Developed context-free grammars in the Java Speech Grammar Format (JSGF) and various other artificial data for ML (seq2seq) model training.
- Worked daily in a Linux environment using Bash, Git, regular expressions, and several other tools and techniques on the command line.

### Linguist

#### Defined AI

 2020    Remote

- Annotated prosody in English utterances using the ToBI guidelines and performed spectrogram analysis using Praat.
- Collaborated with a team of linguists to meet deadlines.

### Assistant to the Laboratory Director


#### Center for Understanding Biology using Imaging Technology

 2018 – 2019    Stony Brook, NY

- Assisted the laboratory director with document formatting, research assistance, academic citations, and record keeping.
- Utilized various software including Microsoft Word, Excel, and EndNote.
- Managed assignments to meet deadlines on a weekly basis.

### Linguist Intern

#### Stony Brook University

 2018    Stony Brook, NY

- Transcribed audio files of endangered languages and annotated utterances using the International Phonetic Alphabet.
- Performed spectrogram analysis and data annotation using Praat.

## EDUCATION


### M.A. in Computational Linguistics

#### Stony Brook University

 Sept. 2019 – January 2021   Cumulative GPA: 3.93

### B.A. in Linguistics, *Summa Cum Laude*

#### Stony Brook University

 Sept. 2016 – Dec. 2018   Cumulative GPA: 3.98

### Coursework in Computer Science

#### Suffolk County College

 Sept. 2013 – May 2016   Cumulative GPA: 4.0

## SKILLS

Python

Linux

Git

GitHub/GitLab

Bash

Regex

Scripting

JSGF

NLP

OCR

Annotation

Data processing

JavaScript

Node

HTML/CSS

Markdown

LaTeX

SQLite

Jira


Confluence

Documentation

## PROJECTS


### News Headline Toolkit

#### Personal Project

-  2024
- Toolkit for creating and interfacing with a database of news headlines.
  - Utilizes Tesseract OCR backend for image-to-text conversion.
  - Includes language modeling, text synthesis, data analysis, and data visualization tools.


### Blog Engine

#### Personal Project

-  2021
- Backend engine for my personal blog, built with Node.
  - Utilizes Node packages Showdown.js and Mustache.js to auto-generate HTML blog posts via Markdown conversion.

### Recommendation Engine

#### Master's Thesis Project

-  2020
- Content-based filtering recommendation engine for academic papers in linguistics from the LingBuzz database.
  - Utilizes similarity matrix based on TF-IDF vectorization values to score and measure similarity between papers.

## LANGUAGES

English   

Japanese   