

# Daniel Galbraith, PhD — Résumé

---

dalgalbraith@gmail.com · (650) 823-6649 · [www.danielgalbraith.com](http://www.danielgalbraith.com) · [github.com/danielgalbraith](https://github.com/danielgalbraith)

## EDUCATION

2018                    **Stanford University**, PhD in Linguistics  
2013                    **University of Cambridge**, MPhil in Linguistics  
2012                    **University of Cambridge**, BA in Modern and Medieval Languages  
                         *First Class Honours with Distinction, ranked first in year*

**Computational Linguist** and Stanford PhD, working in Siri speech synthesis and speech recognition. Spent the last five years dedicated to human language research. Exceptional Data Science and NLP/NLU skills, with five years of experience sourcing, managing and analyzing large quantities of linguistic data. Proven self-starter and highly motivated learner, excited to work on data-driven AI projects and state-of-the-art language models.

*Relevant skills:* Python · R · NLP/NLU (ASR, TTS, PoS Tagging, Dependency Parsing, NER, Language Modeling) · ML/Deep Learning (Keras, TensorFlow, scikit-learn) · Data Science · Statistics · Databases (SQL/PostgreSQL) · Data Visualization (Bokeh) · NLP toolkits (NLTK, CoreNLP, spaCy) · Git · Unix

## PROFESSIONAL EXPERIENCE

**Computational Linguist**, Apple (via Welocalize): June 2018–

- Write daily patches & software updates for Siri speech recognition and text-to-speech
- Design and contribute speech synthesis rulesets, manage pipeline of error logs for self and colleagues
- Perform regular reviews of colleagues' code, personally reduced bug count by 300+ within a month
- Manage own workflow prioritization for frequent deadlines, constantly balance varying priority levels

**Localization QA**, Lionbridge: May–Aug 2017

- Screened UK English sentences for training data, verified phonetic IPA transcription against audio files
- Ran QA tests on gesture functionality, suggestions and autocorrections for Faroese Android virtual keyboard

---

## OTHER RELEVANT EXPERIENCE

- **Completed Data Scientist track with certification** in DataCamp Python specialization (22 courses), including data wrangling, statistical thinking, supervised & unsupervised learning, deep learning and data visualization
- **US Presidential Speeches machine learning** project: analysed & provided gold-standard annotations for sentence stress (academic paper under review), developing model using Keras for automatic stress prediction
- **Trained dependency parser and PoS tagger** with CoreNLP & spaCy CNN model to recognize new tokens; early contributor to test data and syntactic analyses for Stanford/Universal Dependencies
- **Extensive statistical modeling experience** in PhD thesis: predicted grammatical variants via logistic regression with R and Python scikit-learn, constructed own 2.5 million word Faroese blog corpus for training data

---

## COMPUTATIONAL SKILLS

- Built creative **NLP side-projects and linguistic research tools**:
  - **LangGen**: grabs online language statistics & simulates random grammar weighted by frequencies
  - **StressParse**: auto-annotates Faroese-language texts for metrical stress
- **Unix & Linux**: Ubuntu user for 9 years, installed several distributions in various languages
- **Command line tools & shell scripting**: carry out daily tasks using git, grep, vim and bash

## LANGUAGES

- **English** (native), **Faroese, French & Spanish** (university-level), in-depth academic knowledge of multiple others including Icelandic, Danish, Norwegian, Lithuanian, Hebrew, Korean, Mandarin, Turkish & Hungarian