Daniel Galbraith — Résumé

dalgalbraith@gmail.com · (650) 823-6649 · www.danielgalbraith.com · github.com/danielgalbraith

Education

2018 **Stanford University**, PhD in Linguistics
methods: data analysis, statistical modeling, scripting, experimental design, corpora
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 recent Stanford PhD graduate. Spent the last five years sourcing, managing and annotating large quantities of linguistic data. Exceptional skills in data wrangling, acquiring and cleaning training data, text processing, statistical modeling, Python & R scripting, Unix command line, NLP toolkits and annotation schemas. Industry experience in Siri speech localization, TTS & ASR with daily git workflow. Proven self-starter and highly motivated learner.

Professional experience

Computational Linguist, Apple (via Welocalize): 2018–

- Write daily patches & software updates for Siri speech recognition and text-to-speech
- Manage pipeline of error logs for self and colleagues, design and contribute large rulesets for bug fixes
- 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

Research experience

- **Universal Dependencies** NLP project: early contributor to syntactic analyses for Stanford dependency parser, currently advising and providing data pipeline for Faroese syntax
- **US Presidential Speeches** research project: analysed & annotated inaugural addresses for sentence-level stress, advised annotator & wrote documentation detailing stress-marking algorithm
- **Quantitative methods** for PhD thesis and side-projects:
  - extensive **statistical modeling and analysis using R and Python** in thesis, going far beyond academic requirements; committee advised to publish thesis as syntax textbook
  - constructed own **2.5 million-word Faroese blog corpus** using wget and XPath for web-scraping (accessible at danielgalbraith.com/corpora); modeled data via logistic regression with **Python pandas & scikit-learn**
  - designed complex surveys for **eliciting multiple kinds of language data** in fieldwork trips on Faroe Islands and Iceland: textual, audio-visual, native speaker annotations & scalar judgements on sentences

Computer skills

- **Command line tools & shell scripting**: carry out daily tasks using git, grep, vim and bash
- **Python and R scripting** for personal side-projects, including:
  - **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
- **SQL & databases**: completed online SQL tutorials, frequently conduct complex queries of error log database

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