

Daniel Galbraith — Résumé

dagalb@stanford.edu · (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. Outstanding background in data-driven research. Spent the last five years collecting, managing and analyzing large amounts of linguistic data. Exceptional skills in data wrangling, statistical modeling, Python and R scripting, NLTK and annotation schemas. Industry experience in TTS, ASR and speech localization with daily git workflow. Proven self-starter and highly motivated learner.

PROFESSIONAL EXPERIENCE

Computational Linguist, Apple (via Welocalize): 2018–

Write daily patches for Siri ASR and TTS, manage pipeline of error logs for self and colleagues. Constant use of git and Unix command line to provide software updates.

- Design and contribute large rulesets for bug fixes, regularly search and edit large patch files with regex & vim
- Daily reviewer of colleagues' code, extensive teamwork & version control experience
- Manage own workflow prioritization for frequent deadlines, constantly balance varying priority levels
- Provide own self-motivated solutions for recurring problems, going beyond expectations

Localization QA, Lionbridge: May–Aug 2017

- **UK English audio**: Verified English sentences against phonetic IPA transcription and native speaker audio
- **Faroese Android**: Tested gesture functionality, suggestions and autocorrections for Faroese 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 undergraduate annotator & wrote documentation detailing stress-marking algorithm
- **Quantitative methods** for PhD thesis and side-projects:
 - extensive **statistical modeling and analysis** 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 (raw text & HTML accessible at danielgalbraith.com/corpora)
 - designed complex surveys for **eliciting multiple kinds of language data** in fieldwork trips on Faroe Islands and Iceland, analyzed large volumes of native speaker data

COMPUTER SKILLS

- **Python and R scripting** for personal side-projects, including:
 - **LangGen**: generates random grammar weighted by frequencies of feature settings, grabs statistics from World Atlas of Language Structures, currently working to implement grammatical tendencies
 - **WordGen**: generates a given number of word forms based on a predefined phonology
- **Unix & Linux**: Ubuntu user for 9 years, installed several distributions in various languages; carry out daily tasks using command line tools & bash

LANGUAGES

- **Faroese** (fluent), **French & Spanish** (strong competence in both), in-depth academic knowledge of multiple others including Icelandic, Danish, Norwegian, Lithuanian, Hebrew, Korean, Mandarin, Turkish & Hungarian