Daniel Galbraith, PhD

NLP Scientist

3178 Fallen Leaf Street, Palo Alto, CA 94303
650.823.6649
danielg@mail.fo
www.danielgalbraith.com

Education

Stanford University / PhD in Linguistics SEP 2013 - JUN 2018
Research foci: syntax, phonology, NLP · Advisors: Paul Kiparsky, Annie Zaenen, Vera Gribanova, Boris Harizanov
Collaborated on Stanford Dependencies parser with Chris Manning & Tim Dozat, cited 110+ times
University of Cambridge / MPhil in Linguistics OCT 2012 - JUN 2013
University of Cambridge / BA in Modern and Medieval Languages OCT 2008 - JUN 2012
French & Spanish, general linguistics · *First Class Honours with Distinction:* ***Ranked first in class***

Experience

NLP Scientist / Mosaix AI
NOV 2018 - PRESENT, PALO ALTO

Research & development of SoTA NLP system for multilingual voice assistance.

Entity Recognition

- Designed & added semantic annotation logic for training pipeline in multiple domains, achieving entity validation accuracy +10%.
- Developed flow of parsing in Scala codebase through smarter data representation, leading to better end-to-end precision.
- Built a benchmark framework by designing domain-specific metrics for NER and intent classification.

ASR

- Prototyped in-house ASR for English and Swahili voice queries, analyzed quality in notebook & fine-tuned DeepSpeech model.
- Attained large reduction of word error rate (10% en, 28% sw) by combining public acoustic model & in-house language model.

Multi-language Quality

- In charge of Hindi model quality; added both deterministic and ML models for Hindi & Bengali (Bangla).
- Identify needed changes & maintain quality across multiple languages (en, ar, hi, bn, es, vi) using linguistic & ML expertise.
- Created Bengali stemmer from scratch, applying own research in case-marking languages, leading to >20% quality increase.

NLP Research

- Investigating knowledge transfer of BERT attention mechanisms & relation to constituency parse.
- Published single-authored linguistics paper in Cambridge journal on prosodic structure, potential for application to ASR/TTS.
- Experimented with SOTA approaches to syntactic/semantic parsing, integrating insights from linking theory.

Interpretation Ranking

- Determined need for better featurization in ranking/scoring model & added engineered features, e.g. Levenshtein distance, yielding +8% test AUC.
- Tested bias features for specific intents, resulting in >2% error rate reduction for entertainment queries.

Development of ontology/data pipeline

- Guided PMs who manage semantic labelers & voice data consultants, ensuring standards are met for ML pipeline.
- Improved design of ontology for deep parsing and intent classification models using knowledge of semantic categories.

Computational Linguist / Apple (via Welocalize, Inc.)
JUL 2018 - OCT 2018, CUPERTINO

Full-time contractor providing software updates for Siri speech recognition and speech synthesis.

- Wrote daily patches for ASR & TTS improvements.
- Personally reduced bug count by 300+ within first month.

Graduate Research Assistant and Teaching Assistant / Stanford University
JAN 2014 - SEP 2017, STANFORD

Doctoral linguistics research on syntax, metrical phonology and NLP with advisor Paul Kiparsky.
Thesis downloaded 640+ times.

- Proposed new approach to syntactic theory, book draft under review by Oxford University Press.
- Researched factors determining sentence stress on US Presidential speeches project, preprint already cited.
- Early contributor to Stanford Dependencies with Chris Manning, named author on original software.

Skills

Python · Scala · R · NLP & NLU (syntactic/semantic parsing, entity recognition, ASR, TTS) · data science & statistics · hypothesis testing · ML/deep learning (Keras, TensorFlow, scikit-learn) · databases (SQL) · NLP toolkits (NLTK, CoreNLP, spaCy) · Git · Unix · LaTeX · l10n & i18n

Publications

Nivre, Joakim, [...], **Daniel Galbraith**, [...] et al. 2015. *Universal Dependencies 1.2*. Universal Dependencies Consortium. [Also named author of version 2.2, 2018.]

Anttila, Arto, Timothy Dozat, **Daniel Galbraith**, and Naomi Shapiro. 2020. Sentence stress in presidential speeches. To appear in Kentner, Gerrit and Joost Kremers (eds.), Prosody in Syntactic Encoding, special issue of *Linguistische Arbeiten*.

Galbraith, Daniel. 2019. Meter, prosody and performance: evidence from the Faroese ballads. *Nordic Journal of Linguistics* 42(3).227–261.

Galbraith, Daniel. 2018. *The Predictable Case of Faroese*. PhD dissertation, Stanford University. Accessible at <https://purl.stanford.edu/ny726mr8080>.

Galbraith, Daniel. 2013. *Positional and Morphological Case in Faroese*. MPhil thesis, University of Cambridge.

Certification

DataCamp: Data Scientist with Python full track

Languages

Native: English
University-level: French, Spanish, Faroese

International Conferences

A constraint-based account of Faroese ballad meter. 2016. *NordMetrik: Versification, Metrics in Practice*, University of Helsinki.

Faroese ballad meter: a constraint-based approach. 2016. *The 90th Annual Meeting of the Linguistic Society of America*, Washington, DC.

Honours and Awards

Approved National Interest Waiver (merit-based) for US permanent residence

Roberta Bowman Denning Fund for Humanities and Technology with Prof Arto Anttila for US presidential speeches project · George H. Mayr Foundation Graduate Fellowship, Stanford University

Stanford Linguistics departmental and advisor funding awarded for dissertation research conducted on the Faroe Islands and Iceland · Travel scholarship awarded by University of the Faroe Islands

Whalley Prize for highest distinction in final year examinations, University of Cambridge · St. Catharine's College, Cambridge scholarships for First Class results