

# CHRISTOPHER R. HABERLAND

(M) +1 (703) 888-6643 | [haberc@uw.edu](mailto:haberc@uw.edu)

## EDUCATION

**UNIVERSITY OF WASHINGTON** | SEATTLE, WASHINGTON

**Doctor of Philosophy: Linguistics (Computational track)** | Ongoing

**Master of Science: Computational Linguistics** | 2019

Arabic FLAS Fellow, Middle East Center – 2017-18 Academic Year

**UNIVERSITY OF VIRGINIA** | CHARLOTTESVILLE, VIRGINIA | 2015

**Master of Public Policy**

**Concentration: Environmental Policy** | Advisor: William Shobe

**UNIVERSITY OF VIRGINIA** | CHARLOTTESVILLE, VIRGINIA | 2013

**Bachelor of Arts: Double Major in Spanish and Middle Eastern Languages and Literatures**

**Minor: Economics**

Phi Beta Kappa, Golden Key International Honor Society, National Society of Collegiate Scholars, 2013 Critical Language Scholarship - Arabic

## **PUBLICATIONS**

- Haberland, C., & Lao, N. (2023). Kabyle ASR Phonological Error and Network Analysis. Forthcoming, In M. Abbas (Ed.), *Analysis and Application of Natural Language and Speech Processing* (1st ed.). Springer Cham.
- Haberland, C., & Lao, N. (2021). Orthographic Transliteration for Kabyle Speech Recognition. In Proceedings of The Fourth International Conference on Natural Language and Speech Processing (ICNLSP 2021) (pp. 21-32).
- Warziniack, T., Haight, R. G., Yemshanov, D., Apriesnig, J. L., Holmes, T. P., Countryman, A. M., ... & Haberland, C. (2021). Economics of invasive species. *Invasive Species in Forests and Rangelands of the United States*, 305.
- Chen, S. A., Escay, A., Haberland, C., Schneider, T., Staneva, V., & Choe, Y. (2018). Benchmark Dataset for Automatic Damaged Building Detection from Post-Hurricane Remotely Sensed Imagery. *arXiv preprint arXiv:1812.05581*.
- Haberland, C., & Marston, J. (2017, February). Spatial Modeling of Timber Ecosystem Services: Linking the FVS Econ Extension and FSveg Spatial Data Analyzer to Map Stumpage Value. In 2017 Forest Vegetation Simulator (FVS) e-Conference (p. 164).
- Haberland, Christopher. 2017. Mapping Ecosystem Services on the Monongahela National Forest. U.S. Department of Agriculture, Forest Service, Research and Development.

## **LANGUAGES**

**Native:** English | **Advanced:** Spanish | **Intermediate:** Italian, Arabic, Western Persian |

**Beginner:** Ligurian, Basque | **Analytical:** Coahuilteco, P'urhépecha

## SKILLS

**Programming:** Python, Scala | **Database:** Redis, MongoDB | **Cloud:** Google Cloud Platform, AWS  
| **Geographical:** ArcGIS

## PROFESSIONAL EXPERIENCE

**USAA, ADVANCED ANALYTICS AND MACHINE LEARNING | SEATTLE, WA**  
**NLP DATA SCIENTIST | JULY 2020 – PRESENT**

### Information Extraction (IE) Modeling and Ontology Design

- Design and implement patent-pending IE model and associated ontologization process for identifying interpretable cause-of-loss features from millions of unstructured text notes to qualify and quantify property claims and loss characteristics
- Develop NLP pipeline consisting of named entity recognition (NER), entity typing, and semantic role labeling components to create a novel graph-based semantic index of property and auto claims incorporating modality and negation
- Design sliding-window document structure and form parser for claims narrative texts to record subject-complement relations between entities and contextual headers of notes to enable search and filtering using logical queries
- Supervise and evaluate models for risk assessment, subrogation determination, and operational goals using NLP-derived features that prevent millions of dollars in losses

### Conversational Modeling and Transcript Analysis

- Clean, analyze, implement, and optimize thematic intent and empathy model pipelines to batch-classify thousands of call center dialogues per day
- Fine-tune and deploy foundational language models for summarizing root-causes of complaints reported in agent-member dialogues
- Statistically model and compare features between conversational text corpuses to reveal variables correlative with negative events

### NLP Product Research

- Placed 1st out of a field of 250 colleagues in the company-wide Centennial Hackathon after leading the development of an image-fraud fighting solution on team of 5
- Identify and collaborate with in-company stakeholders (Claims Innovation, Property and Innovation) to develop analytical solutions using traditional and NLP-derived data sources
- Collaborate and conduct code review with junior Data Scientists and Data Analysts to ensure adoption of quality software engineering principles and efficient model development
- Present and share new research and best practices from conferences in the NLP field with colleagues after attending NAACL, EMNLP, OpenSource Summit
- Placed second out of 25 in 2022 week-long company-wide hackathon for ideating novel product solution to streamline property underwriting flow by utilizing coupled image-text foundational models on user-submitted photos
- Independently published on modeling unit interventions of different orthographies for downstream neural speech-to-text model performance and presented work at ICNLSP 2021, to appear in the forthcoming journal series *Signals and Communication Technology*

Mosaix.ai | PALO ALTO, CA

NLP ENGINEER | SEPTEMBER 2018 – JUNE 2020

### Multilingual Quality Improvement

- Responsible for all quality improvements related to Arabic annotation, intent classification, and entity recognition models for both dialectal and standard forms of the language
- Wrote deterministic linguistic rules and conducted sensitivity analyses of normalization and tokenization procedures for several languages (English, Arabic, Spanish, Bengali, Hindi, Vietnamese) to improve downstream top-K retrieval performance by more than 10%
- Incorporated language identification and spell-checking functionalities to API

### Entity Recognition and Intent Classification

- Refined ML model logic to achieve entity recognition F1 improvements of over 20% for Arabic pipeline
- Coded efficient stratified sampling procedures for diverse training data augmentation
- Created Git-mining visualization tool to track model quality over time via Git Hooks

### Ontology Design

- Reviewed and redesigned ontology for multilingual voice assistant and semantic search to optimize for customer needs and improve intent inference precision

### Data Pipelining

- Designed and carried out data collection projects with custom-built labeling tools to support natural language understanding (NLU) model training
- Architected a custom data collection pipeline with NoSQL/MongoDB databases to store semantic annotations
- Enforced database schemas for data standardization and interoperability between multiple product components
- Identified and incorporated sources of open data (Wikidata, Anghami) for construction of a knowledge graph (KG) for entity recognition
- Designed a collaborative spreadsheet system with advanced macros for managing a team of annotators and reviewing their annotations
- Designed and managed systems for analyzing company data and created distributable graphical user interfaces (GUIs) for data analysis to support non-technical staff

### Research

- Investigated character-based BiLSTM-CRF Arabic text segmentation for improved multi-dialectal system performance
- Prototyped DeepSpeech automatic speech recognition (ASR) system for Arabic and voice queries and iteratively modified model parameters and preprocessing schemata to achieve a character-error-rate of ~6% on multi-dialectal Arabic

**University of Washington | SEATTLE, WA**  
**RESEARCH ASSISTANT | MARCH 2018 – AUGUST 2019**

DARPA LORELEI project

Lead Investigator: Dr. Gina-Anne Levow

- Designed and developed optical character recognition pipeline to extract multilingual data from printed materials for language modeling of low-resource languages using Tesseract OCR, fastText, and other open source software

Data Science for Social Good Fellow

Lead Investigator: Dr. Youngjun Choe

- Created data pipeline to automatically identify post-hurricane damages from remotely-sensed images by implementing a single-shot multibox detector algorithm with Tensorflow
- Analyzed model performance and tuned hyper-parameters of neural models to improve mean average precision of object detection on custom development data sets

Forest Service ChatBot for Conversational Data Solicitation

Lead Investigator: Dr. Spencer Wood

- Designed and deployed a conversational chatbot to distribute information to and collect input from visitors to National Forests

Topic Modeling and Classification of TripAdvisor Reports

Lead Investigator: Dr. Spencer Wood

- Designed the ontological tagging schema and annotation program to create gold-label training data of social media documents from tripadvisor.com
- Evaluated skip-gram, cbow, SVM, logistic regression, and random forest models for classification of tripadvisor.com text data that evinces economic activity dependent upon coral reefs

**U.S. Forest Service | WASHINGTON, D.C.**  
**FELLOW | OCTOBER 2015 – AUGUST 2017**

Geospatial Application for Analyzing Economic Trade-offs across Landscapes

- Developed web-based GIS applications that facilitate the geospatial identification of economic values of resources on Forest Service lands
- Architected an online library of temporal and geospatial data and created visualizations of economic trade-offs across landscapes using ArcGIS, QGIS software, R, Python, and Excel

Geospatial Modeling of the Economic Value of Forest Stand Aesthetics

- Synthesized environmental-economic literature on forest with U.S. Forest Service data to

create a spatially-explicit scale of relative values of forest stand aesthetic values for the Monongahela National Forest

#### Launch the National Center for Natural Resource Economics Research (NCNRER)

- Organized and documented extant Agency forest economist research expertise by using topic modeling and semantic clustering of scientists' research articles
- Designed and maintained an online clearinghouse of geospatial tools and methodologies for ecosystem services valuation and provided comprehensive software utility evaluation for senior research scientists and policy analysts within the agency

#### Design Urban Tree Shade Index Mobile App

- Led team of four programmers to create GIS-enabled Shade Index mobile app to estimate shade provided by trees in Washington D.C. using open-source municipal data

**University of Virginia | CHARLOTTESVILLE, VA**  
**RESEARCH ASSISTANT | JANUARY 2014 – SEPTEMBER 2015**

#### Research on the Virginia Energy Plan

- Evaluated the assertions contained in a state-commissioned study on the projected economic impacts of the Clean Power Plan
- Organized meetings with energy utility employees, state government officials, and other interested stakeholders to obtain data and exchange relevant information, and compiled and maintained datasets using Excel and R
- Developed static models to project future energy sector trajectories and estimate the feasibility of different pathways

#### Compliance with the EPA's Clean Power Plan: Virginia's Challenges

- Conducted an independent policy analysis on the optimal way for Virginia to respond to the EPA's Clean Power Plan regulations on CO<sub>2</sub> emissions from the electricity sector

**U.S. Department of State | CHARLOTTESVILLE, VA**  
**INTERN | OCTOBER 2013 – AUGUST 2017**

#### "ConSoul" Chatbot

- Managed a team of 2 junior developers to design a chatbot that answers questions about foreign countries and their bilateral relations with the United States using a combination of platforms and tools, including Heroku, api.ai, Python, and MongoDB.

#### Analysis of Adaptation Projects in Developing Countries

- Designed a rule-based system to systematically survey the prevalence of the language of incentives in several countries' climate policies by locating syntactic construction evincing strong commitments to enacting specific policies
- Conducted study that concluded that developing countries' policies often lack concrete measures to incentivize public and private actors to adapt to climate change