

Emily Soward

(541) 543-4215 • eksoward@gmail.com • [linkedin.com/in/eksoward](https://www.linkedin.com/in/eksoward) • github.com/eksoward

I am an experienced NLP data scientist currently seeking opportunities in data R&D. I specialize in reverse engineering, documentation, technical product representation, end-to-end optimization, novel feature deployment, and long-term R&D visioning and task prioritization for in-production enterprise and legacy AI systems.

Skills

Top Skills: Python, C++/C, UNIX, Linux, Bash, regular expressions, Agile development management
Data: Praat, Festival, HTK, HTS, NLTK, MATLAB, R, Octave, Java Drools
Build: CMAKE, CI, baseline and regression test development, Ant, Eclipse IDE, file versioning
Formats: XML, HTML, CSS3, NoSQL, MySQL, Apache Lucene & Solr, JSON, \LaTeX , UML, Doxygen
Basic Skills: ActionScript (including off-label uses), JavaScript/JQuery, KornShell, Scheme, Java, Ruby

Employment

- Nov 2015 – Mar 2017 **Data Scientist — Rosetta Stone Speech Recognition Engineering**
- Co-developed first third-party speech recognition enabled app on the Apple Watch
 - Owned end-to-end product development and publication of acoustic data modeling and augmentation strategy for low-resource speech recognition supporting the Chickasaw Nation
 - Evangelized and consulted on our SDK in developer-centric internal lecture series
 - Lead technical review regarding patent and IP security for data-augmented products
 - Consulted on product feasibility and development for internal teams and external partners
 - Expanded CI testing to baseline recognition performance on real-world test scenarios
 - Strategist and owner on long term speculative speech research projects in K-12
- Jun 2014 – Nov 2014 **Research Linguist II — UnitedHealth Optum360 NLP Innovation Lab**
- Trained and transitioned multi-stage medical coding engines and semantic reasoning systems from ICD-9 to ICD-10 standards (5X complexity increase) for specialty medicine
 - Co-founded cross-project initiative to develop Solr-backed medical text analysis tools for improving research workflows
 - Liaised with hospitals and human coders to solicit UX feedback in early-adopter program
- Feb 2014 – Jun 2014 **OneLink Engineer — TransPerfect GlobalLink**
- Tuned custom web crawlers to scrape and parse language data using in-house tools
 - Delivered on rapid client support turn-arounds and kept pace with work for multiple clients and project teams simultaneously

Graduate Education

- 2012–2013 **Master of Science: Informatics, AI & NLP w. Distinction — University of Edinburgh Centre For Speech Technology Research (CSTR)**
Thesis: *Conversational Speech Corpus Design and Recording for HMM-Based TTS*
- Edinburgh Informatics International Master's Scholar
 - Postgraduate Representative, Informatics Academic Council, College of Science and Engineering Teaching Committee, Edinburgh Senatus Academicus
- Courses:** Advanced NLP, Speech Synthesis, Automatic Speech Recognition, Speech Processing, Natural Language Generation, Machine Translation

Undergraduate Education

- 2010–2012 **Bachelor of Arts: Linguistics — University of Oregon**
Focus on theoretical functional linguistics, language dataset analysis, and computation.
- Research Assistant, Perceptual Phonetics and Nahuatl Language Dictionary Project
 - Founding Editorial Board, Oregon Undergraduate Research Journal
- Courses: Corpus Linguistics, Phonetics, Phonology, Morphology & Syntax I–III, Language and Cognition, Semantics, Discourse Structures, Discrete Math I–III**
- 2006–2010 **Bachelor of Arts: Medieval Studies — University of Oregon**
Focus on language spread, usage patterns, writing, and translation
- U.S. State Dept. Benjamin A. Gilman Scholar, Norway 2009–2010
 - President-Elect, International Students' Union, Univ. Oslo. 2009–2010

Professional Service

- 2015–2017 **IEEE International Region 5 Executive Board, Educational Activities Chairperson**
IEEE USA Educational Activities Board Member

Conference Proceedings and Talks

- May 2018 **General-Session Talk: Write The Docs Portland (Proposal Pending)**
"Document up your Bus Factor: Non-tool techniques for seeking out, extracting, and recording critical information after the bus hits"
General-session talk aimed at describing specific strategies for recovering ecosystem and architecture information needed to document complex legacy software that is partially or completely undocumented
- April 2017 **Invited Talk: SpeechTEK**
"Speech Technology for Augmenting Language Learning Experiences"
Discussion of practical issues and challenges in using speech technologies in language learning software, breaking down oral language instruction into machine-solvable tasks, and using AI to generate appropriate learner feedback
- March 2017 **Proceedings: Int'l Conference on Language Documentation and Conservation**
"Challenges in creating speech recognition for endangered language CALL: A Chickasaw case study"
Case study overview of strategies and optimizations developed to overcome data sparsity and human-factor challenges for acoustic models used in endangered language ASR