

Sean Simpson

Computational Sociolinguist

Contact

sss237@georgetown.edu
seanskylersimpson.com

Natural Languages

English (Native)
Portuguese (advanced)
French (intermediate)

Programming Languages

Python
R
SQL
Bash
HTML5
CSS3
LaTeX

ML Tools

TensorFlow
Keras
Scikit-Learn
SciPy
NumPy
Pandas
Caret

NLP Tools

SpaCy
OpenNLP
Stanford CoreNLP
Gensim
NLTK
TextmineR

EDUCATION

2019 (expected)	Doctorate of Philosophy Computational Linguistics	Georgetown University
2016	Master of Science Sociolinguistics	Georgetown University
2013	Master of Arts Language Documentation & Conservation	University of Hawaii at Manoa
2011	Bachelor of Arts Linguistics, Anthropology	University of Florida

Work Experience

2018 (Aug-Nov)	Google <i>Analytical Linguist Intern, Ads Human Eval</i> Developed machine learning models to predict inter-rater reliability and task completion time for new survey templates designed by the Ads Human Evaluation team. Developed an internal R package to aid in the extraction of structural, semantic, and lexical features from survey templates. Integrated these machine learning models and the feature extraction package into the post-pilot analysis pipeline to determine whether or not to proceed with new Human Eval survey templates.	Los Angeles, CA
2018 (May-Aug)	International Monetary Fund (IMF) <i>Projects Officer, Innovation Lab Unit</i> Provided consultation and development expertise on a number of issues related to Natural Language Processing and Machine Learning. Primarily focused on designing and implementing a system to predict financial crises based on natural language features extracted from news corpora.	Washington, DC

2016 - 2017	<p>Center for Advanced Study of Language (CASL) <i>NLP Research Specialist</i> Designed and implemented a Natural Language Processing system to detect newly emerging drug terminology in streaming Twitter corpora. This detection system successfully identified a number of drug terms previously unknown to subject matter experts from the Center for Substance Abuse Research as they emerged online in real time over the course of several months.</p>	College Park, MD
2015	<p>Center for Advanced Study of Language (CASL) <i>Research Assistant</i> Managed the collection of a tri-lingual SMS corpus for the purpose of cross-language variation analysis.</p>	College Park, MD
2012 - 2013	<p>Catalogue of Endangered Languages (ELCat) <i>Project Coordinator</i> Lead a team of research assistants in collecting and assembling the language data for the 3,000+ languages that appear on the Endangered Languages Project website (WWW.endangeredlanguages.com)</p>	Honolulu, HI
2011 - 2012	<p>Language Documentation Training Center (LDTC) <i>Director</i> Managed the day-to-day necessities of the training center and taught weekly workshops on issues of language documentation to members of minority and threatened language communities.</p>	Honolulu, HI

PUBLICATIONS

Peer-Reviewed Journal Articles

2018	<p>Simpson, S. Sean, Nikki Adams, M. Claudia Brugman, and J. Thomas Conners. Detecting Novel and Emerging Drug Terms Using Natural Language Processing: A Social Media Corpus Study. <i>Journal of Medical Internet Research: Public Health and Surveillance</i>, 4(1). e2</p>
2016	<p>Kirtley, Joelle, Katie Drager, James Grama, and Sean Simpson. An acoustic analysis of the vowels of Hawai'i English. <i>Journal of the International Phonetic Association</i>, 46(1). 79–97</p>
2013	<p>Drager, Katie, Joelle Kirtley, James Grama, and Sean Simpson. Language variation and change in Hawai'i English: KIT, DRESS, and TRAP. <i>University of Pennsylvania Working Papers in Linguistics</i>, 19(2). 40–50</p>
2013	<p>Simpson, Sean. The Endangered Languages Project: A digital tool for research and engagement. <i>Katoba to Shakai (Language and Society)</i>, 15. 135–154</p>

Peer-Reviewed Conference Presentations

- 2017 Simpson, Sean. Reconstructing Thread Structure For Reddit Comments. Poster presented at *Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL)*. Washington, DC
- 2016 Simpson, Sean. Keeping up with Cait: A longitudinal analysis of F0 and [s] in the speech of Caitlyn Jenner. Poster presented at *New Ways of Analyzing Variation (NWA) 45*. Vancouver, BC
- 2014 Drager, Katie, Sean Simpson, James Grama, and Joelle Kirtley. Variation and change in the back vowels of Hawai'i English. Paper presented at *New Ways of Analyzing Variation (NWA) 43*. Chicago, IL
- 2014 Simpson, Sean. Phonetic cues to Localness in Hawai'i English: Feature clusters influencing perception. Paper presented at *New Ways of Analyzing Variation (NWA) 43*. Chicago, IL
- 2014 Simpson, Sean. The Endangered Languages Project: Overview and assessment. Paper presented at *The LSA Annual Meeting*. Minneapolis, MN
- 2013 Campbell, Lyle, Nala Lee, Eve Okura, Sean Simpson, and Kaori Ueki. New knowledge: Findings from the Catalogue of Endangered Languages. Paper presented at *The 3rd International Conference on Language Documentation and Conservation (ICLDC)*. Honolulu, HI
- 2013 Simpson, Sean. Practical applications of the Endangered Languages Project. Paper presented at *The 3rd International Conference on Language Documentation and Conservation (ICLDC)*. Honolulu, HI
- 2013 Simpson, Sean. Preliminary results: Phonetic variation and perceived Localness in Hawai'i. Paper presented at *The 17th Annual Graduate Student Conference of Languages, Linguistics and Literature*. Honolulu, HI
- 2012 Grama, James, Joelle Kirtley, Katie Drager, Sean Simpson, and Rebecca Clifford. Language variation and change in Hawai'i English: KIT, DRESS and TRAP in motion. Paper presented at *New Ways of Analyzing Variation (NWA) 41*. Bloomington, IN

Book Chapters

- 2018 Belew, Anna and Sean Simpson. Language extinction then and now. In Campbell, Lyle and Anna Belew (eds.). *Cataloguing the Endangered Languages of the World*. Abingdon: Taylor and Francis
- 2018 Belew, Anna and Sean Simpson. The status of the world's endangered languages. In Campbell, Lyle and Kenneth Rehg (eds.). *The Oxford Handbook of Endangered Languages*. Oxford: Oxford University Press
- 2018 Heaton, Raina and Sean Simpson. How ELCat serves communities whose languages are at risk. In Campbell, Lyle, John Van Way, and Anna Belew (eds.). *Cataloguing the Endangered Languages of the World*. Abingdon: Taylor and Francis

Teaching Experience

2018, spring	Teaching Assistant LING/COSC-572: Empirical Methods in NLP	Georgetown University
2017, fall	Teaching Assistant LING-367: Computational Corpus Linguistics	Georgetown University
2017, spring	Teaching Assistant LING-469: Analyzing Language Data with R	Georgetown University
2016, fall	Teaching Assistant LING-481: Sociolinguistic Variation	Georgetown University
2015, fall	Instructor LING-001: Introduction to Language	Georgetown University
2015, spring	Instructor LING-001: Introduction to Language	Georgetown University
2014, fall	Instructor LING-001: Introduction to Language	Georgetown University