Andrew Zupon
The University of Arizona

✉ zupon@email.arizona.edu
🖰 zupon.github.io

Education

02022	**PhD in Linguistics** (anticipated)
	The University of Arizona
02019	**MS in Human Language Technology**
	The University of Arizona
02015	**MA in Linguistics**
	The University of Utah
02013	**Honors BA in Linguistics (Major), Spanish (Major), Russian (minor)**
	The University of Utah

Non-Degree Programs

02019	**Breton Linguistic and Cultural Heritage Studies Summer School**
	University of Western Brittany, Kemper, France
02012	**Russian Language and Culture Intensive Summer Program**
	Krasnoyarsk State Pedagogical University, Krasnoyarsk, Russia
02011	**Spanish History and Culture Intensive Summer Program**
	University of Oviedo, Xixón, Spain

Employment

02020–present	**NLP Engineer Consultant**, Lum AI (Tucson, AZ)
	My contributions include writing grammars for extracting new types of causal arguments, developing an evaluation framework for entity grounding, and creating unit tests for new features, all for the DARPA Causal Exploration project.
Summer 02020	**Linguist Intern**, Google, Inc. (remote; team in New York City, NY)
	I built an open-source text normalization pipeline using Pynini to evaluate different approaches to normalization as well as the quality of existing data sets for a selection of low-resource languages of Africa. `https://github.com/googleinterns/text-norm-for-low-resource-languages`
Summer 02018	**Chatbot Intern**, Nextiva (Scottsdale, AZ)
	I used the Rasa toolkit to help develop a company-internal chatbot. Tasks included creating training data for the chatbot's language model and dialog model, running and comparing different models, and experimenting with new components for the chatbot.
02015–16	**Junior Linguist**, Adecco at Google (Mountain View, CA)
	I performed linguistic annotation, corrections, and data analysis for various NLP projects, including syntactic parsing, part-of-speech tagging, named entity recognition (NER), coreference resolution, word sense disambiguation, and text summarization.

Research Experience

02018–present	**Research Assistant**, Dr. Mihai Surdeanu
	Department of Computer Science, University of Arizona
	Worked on implementing interpretable models for relation extraction. Wrote grammars for event extraction from text (biomedical relations and human migration events). Currently working on building a compositional ontology for causal events and developing the infrastructure for using the ontology.
Fall 02018	**Research Assistant**, Dr. Andrew Carnie
	Department of Linguistics, University of Arizona
	Prepared overview and project descriptions for an NSF grant on Breton initial consonant mutations.
02012–15	**Research Assistant**, Dr. Rachel Hayes-Harb, Dr. Shannon Barrios
	Speech Acquisition Lab, Department of Linguistics, University of Utah
	Ran phonetic/phonological acquisition experiments involving human subjects.

Other Projects

Detecting Corpus Annotation Inconsistencies for Syntactic Parsing

Examples found in one data set that are not found in another can represent unseen examples, inconsistencies in annotation, or annotation errors. Given two data sets, can we detect mismatches in labels between them, use those mismatches to identify annotation irregularities, and automatically correct them?

Scottish Gaelic Part-of-Speech Tagger

I developed a part-of-speech tagger for Scottish Gaelic, a low-resource language, comparing a Hidden Markov model with an LSTM neural network architecture. My LSTM implementation achieves over 80% accuracy on overall tags. Previous work (Lamb and Danso 2014) achieves 76.6% accuracy on the same dataset using a Brill bigram tagger.

Ossetian Language Elicitation and Analysis

I participated in an Ossetian phonology and morphology linguistic elicitation project with the Iranian Languages Research Group at the University of Arizona. This included compiling elicitation materials, conducting interviews with a native speaker, transcribing the audio recordings, and analyzing the data.

Peer-Reviewed Publications

Zupon, Andrew, Faiz Rafique, and Mihai Surdeanu. (02020). An Analysis of Capsule Networks for Part of Speech Tagging in High- and Low-resource Scenarios. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Workshop on Insights from Negative Results in NLP.* 66–70 http://dx.doi.org/10.18653/v1/2020.insights-1.10

Zupon, Andrew, Maria Alexeeva, Marco A. Valenzuela-Escárcega, Ajay Nagesh, and Mihai Surdeanu. (02019). Lightly Supervised Representation Learning with Global Interpretability. *Proceedings of the 3rd Workshop on Structured Prediction for NLP.* 18–28. http://dx.doi.org/10.18653/v1/W19-1504

Conference Presentations

Zupon, Andrew. (02021). "Kashmiri Word Order Variation and the Uniformity of V2 Structure". 26th Himalayan Languages Symposium. Paris, France, September 1–3. **Conference postponed from 02020 to 02021 due to COVID-19 pandemic.**

Zupon, Andrew, Faiz Rafique, and Mihai Surdeanu. (02020). "An Analysis of Capsule Networks for Part of Speech Tagging in High- and Low-resource Scenarios". Workshop on Insights from Negative Results in NLP, Virtual Conference, November 19.

Zupon, Andrew, Maria Alexeeva, Marco A. Valenzuela-Escárcega, Ajay Nagesh, and Mihai Surdeanu. (02019). "Lightly Supervised Representation Learning with Global Interpretability". 3rd Workshop on Structured Prediction for NLP, Minneapolis, MN, June 7.

Zupon, Andrew. (02018). "Tagging Scottish Gaelic". Arizona Linguistics Circle 12 (ALC12), Tucson, AZ, October 12–14.

Zupon, Andrew, Adam King, Colleen Patton, Michael Hammond, and Andrew Carnie. (02018). "A Corpus Analysis of Breton Soft Mutation Inconsistencies". Typologi ar Brezhoneg (Typology of Breton), Quimper, France, June 19–20.

Zupon, Andrew. (02016). "Word Order(*s) in Kashmiri". Arizona Linguistics Circle 10 (ALC10), Tucson, AZ, December 2–4.

Zupon, Andrew. (02015). "Agreement Restrictions in Icelandic Quirky Subject Constructions". Seoul International Conference on Generative Grammar (SICOGG) 17, Seoul, Republic of Korea, August 5–8.

Zupon, Andrew. (02015). "Kashmiri Word Order(*s): V1, V2, V3, V4, and VF". 21st Himalayan Languages Symposium, Kathmandu, Nepal, June 2–3. **Accepted, did not present due to earthquakes and conference postponement.**

Zupon, Andrew. (02015). "Icelandic Quirky Agreement Restrictions". University of Utah Student Conference in Linguistics, Salt Lake City, UT, April 17.

Zupon, Andrew. (02015). "Kashmiri Word Order(*s): V1, V2, V3, V4, and VF". University of Utah Student Conference in Linguistics, Salt Lake City, UT, April 17.

Zupon, Andrew. (02015). "Icelandic Quirky Agreement Restrictions". L'atelier de syntaxe Montréal-Ottawa-Toronto-Hamilton Syntax Workshop (MOTH), Ottawa, Ontario, Canada, March 28.

Zupon, Andrew. (02014). "A Topicalization Analysis of Icelandic Quirky Case Word Order Facts". University of Utah Student Conference in Linguistics, Salt Lake City, UT, April 11.

Zupon, Andrew. (02013). "Restrictions on Denominal Verb Formation". University of Utah Undergraduate Research Symposium, Salt Lake City, UT, April 3, 02013.

Invited Talks

02021   CareerPath Event. Department of Linguistics, University of Utah, March 4.

02017   "Linguistics Beyond Academia". Department of Linguistics, University of Utah, April 13.

Teaching Experience

Primary Instructor

Summer 02018	LING 123 *Introduction to Mathematical Approaches to Language* (UA, online)
Spring 02018	LING 123 *Introduction to Mathematical Approaches to Language* (UA, online)
Spring 02014	ESL 1040 *Grammar and Editing for Non-Native Speakers* (UU, in person)

Graduate Teaching Assistant/Associate

Fall 02017	LING 300 *Introduction to Syntax*, Dr. Simin Karimi (UA, in person)
	LING 150A *Language in the World*, Dr. Amy Fountain (UA, in person)
Spring 02017	LING/PSY/SLHS 341 *Language Development*, Dr. LouAnn Gerken (UA, online)
Fall 02016	LING 150A *Language in the World*, Dr. Amy Fountain (UA, in person)
Spring 02015	LING 1069 *Bad Words and Taboo Terms*, Dr. Randall Eggert (UU, online)
Fall 02014	LING 2010 *Introduction to the Study of Language*, Dr. Nathan Vooge (UU, online)
	LING 1069 *Bad Words and Taboo Terms*, Dr. Randall Eggert (UU, online)

Guest Lectures

Fall 02017	LING 300 *Introduction to Syntax*, Dr. Simin Karimi (UA), three lectures.
Fall 02017	"History of English". LING 150A *Language in the World*, Dr. Amy Fountain (UA), November 29, 02017.

Service & Outreach

02021	**Reviewer**, Coyote Papers (Arizona Linguistics Circle proceedings journal)
02019	**President, Faculty Representative**, LingCircle (UA Linguistics graduate student organization)
02018–19	**Travel Grant Reviewer**, Graduate & Professional Student Center, UA
02018	**Vice President, Treasurer, Social Chair**, LingCircle (UA Linguistics graduate student organization)
02017–20	**Organizing Committee, Abstract Reviewer**, Arizona Linguistics Circle 11–14
02017	**Outreach Volunteer**, Tucson Festival of Books
02015	**Student Member**, Department of Linguistics Research & Community Committee, University of Utah
02015	**US Contest Site Coordinator**, North American Computational Linguistics Olympiad
02015	**Conference Chair**, University of Utah Student Conference in Linguistics
02012–13	**Peer Advisor**, Department of Languages and Literature, University of Utah
02012	**San Juan River Environmental Cleanup**, Alternative Spring Break, University of Utah

Technical Skills

Programming:	Python, Scala, Perl, Prolog, Matlab
Toolkits/Libraries:	DyNet, NLTK, NumPy, spaCy, Pynini, Lucene, Kaldi, Rasa
General:	Linux, Bash/Shell, git, vim, Jupyter Notebook, LaTeX

Language Skills

Native:	English
Intermediate:	Spanish, Russian
Basic:	Breton, French, Mandarin