Maria Alexeeva

University of Arizona

maxaalexeeva@gmail.com
https://maxaalexeeva.github.io/
https://github.com/maxaalexeeva

Summary: I am a PhD candidate in Linguistics and a Research Associate in the Department of Computer Science, focusing on Natural Language Processing and providing technical support and analysis for various DARPA-funded projects.

Research Interests: information extraction, rule-based approaches, natural language generation as a way to augment rule-based approaches, extraction of subjective views from scientific publications

Education

The University of Arizona Tucson, AZ

- Ph.D. in Linguistics (anticipated 2024)
- M.S. in Human Language Technology (2019)

The University of Utah Salt Lake City, UT

- M.A. in Linguistics (2015)
- Graduate Teaching English to Speakers of Other Languages (TESOL) Certificate

Far Eastern University of Humanities Khabarovsk, Russia

- Specialist in Linguistics and Translation, English, and Mandarin Chinese (2010)

Relevant Experience

Graduate Researcher (NLP) | CLU Lab, University of Arizona (August 2018–present)
Supervised by Dr. M. Surdeanu, Dr. R. Sharp, and Dr. C. Morrison

Present

- Serving as a key researcher and developer for the *HEURISTICS* project under the DARPA *Habitus* program, working on neuro-symbolic information extraction systems, with the tasks including data collection, rule-based extraction, model training, and large-language model (LLM) fine-tuning and output evaluation in the agronomy domain
- Developing datasets: defining tasks, preparing annotation guidelines and training materials, collecting annotations using crowdsourcing and in-house annotation, and conducting quality control
- Providing onboarding and mentorship support to junior researchers

Past

- Contributed to the development of a tool for semi-automated expert knowledge curation and ran participant sessions for the associated user study (journal article under review)
- Led the development of an automatic machine reading pipeline for the *AutoMATES* project (part of the DARPA *Automating Scientific Knowledge Extraction (ASKE)* program)
- Translated information extraction rules from Stanford Dependencies to Universal Dependencies for a biomedical information extraction system (REACH)

NLP Engineer Consultant | Lum AI Tucson, AZ (2020–July 2023)

- Worked on a variety of NLP-related tasks, which included proposing solutions to achieve project goals; using, modifying, and evaluating in-house tools; data annotation; and error analysis
- Latest project: rule-based information extraction from spoken interview transcripts in the agronomy domain

NLP Intern | Crane AI New York City, NY (Summer 2018)

- Implemented intent classifiers and trained entity extractors using RASA NLP
- Researched algorithms and tools to solve problems at hand and evaluated their success potential given the resources available
- Gathered, preprocessed, and analyzed data using various NLP toolkits

Technical Experience

Programming:	Python, Scala
Toolkits/Libraries:	Scikit-Learn, NLTK, NumPy, pandas, NetworkX, neo4j, transformers
General:	Linux, MacOS, Bash, Jupyter Notebook, git, Amazon Mechanical Turk

Publications

Alexeeva, Maria, Caroline Campbell Hyland, Keith Alcock, Allegra A. Beal Cohen, Hubert Kanyamahanga, Isaac Kobby Anni, Mihai Surdeanu. "Annotating and Training for Population Subjective Views." *13th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis. 2023.* https://aclanthology.org/2023.wassa-1.36/ (peer-reviewed)

Alexeeva, Maria, Allegra A. Beal Cohen, and Mihai Surdeanu. "Combining Extraction and Generation for Constructing Belief-Consequence Causal Links." *Proceedings of the Third Workshop on Insights from Negative Results in NLP. 2022.* https://aclanthology.org/2022.insights-1.22/ (peer-reviewed)

Alexeeva, Maria, Rebecca Sharp, Marco A. Valenzuela-Escárcega, Jennifer Kadowaki, Adarsh Pyarelal, and Clayton Morrison. "Mathalign: Linking formula identifiers to their contextual natural language descriptions." *In Proceedings of The 12th Language Resources and Evaluation Conference*, pp. 2204-2212. 2020. https://aclanthology.org/2020.lrec-1.269/ (peer-reviewed)

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 (peer-reviewed)

Christison, MaryAnn, and **Maria Alexeeva**. "Formal Characteristics of Learner Language." The TESOL Encyclopedia of English Language Teaching (2018): 1-7.

Other Projects

Impact of Clause Boundary Information on the Performance of a Syntactic Dependency Parser
Using clause boundary information, we significantly improved performance of the head identification component of a syntactic dependency parser compared to the baseline in several conditions.

Analyzing the data from Lifestyle Intervention for Ovarian Cancer Enhanced Survival (LIVES) Study
The work was part of a broader effort to explore the possibility of using machine learning techniques in the process of evaluating motivational counseling calls in terms of their ability to improve lifestyle behavior outcomes of cancer survivors. Given the data for two arms of the study, we analyzed the data for whether or not the two arms are sufficiently distinct to be automatically classified; we also explored the possibility of automatically aligning participants' responses to compare reactions of interviewers to similar utterances between the two arms of the experiment.