Sarah Moeller

Department of Linguistics
Institute of Cognitive Science
Hellems 290, 295 UCB, Boulder, CO
sarah.moeller@colorado.edu
sarahrmoeller.github.io

EDUCATION

Ph.D. Linguistics and Cognitive Science   May 2021
University of Colorado Boulder
> Dissertation: "Integrating Machine Learning into Language Documentation and Description"
> Committee: Mans Hulden, Martha Palmer, Andrew Cowell, Alexis Palmer, Katharina Kann

Certificate in College Teaching   2021
University of Colorado Boulder

M.A. Linguistics   2020
University of Colorado Boulder

M.A. Applied Linguistics   2010
Dallas International University

B.A. History   2002
Thomas Edison State University

RESEARCH PROJECTS

Arapaho Online Dictionary Project   2020-present
21st Century Tools for Indigenous Languages
> Goal: to connect computational morphological parsers to online dictionary to improve usability by users with limited fluency in Arapaho.

Integrating Machine Learning into Language Documentation and Description   2016-present
University of Colorado
> Research question: how does early analytic decisions by linguists inform NLP models and how should integration of machine learning influence linguists' workflow? I devised new NLP methods for morphological analysis and human-in-the-loop description of morphological paradigms and tested assumptions about the importance of part-of-speech (POS) tagging. Publications in *Workshop on Polysynthetic Languages*, *ComputEL*, *EMNLP*.

Universal Meaning Representations (UMR)   2019-2020
Computational Language and Education Research (CLEAR)
> Project aim: develop new cross-linguistic abstract meaning representations. I tested the guidelines on English and Russian. I built a proposition bank of semantic frames for highly-frequent Russian predicates

and labeled semantic roles to create a publicly-available gold standard corpus used for automated English-to-Russian projection of semantic roles. Publication in *LREC*.

"Teacher-in-the-Loop"   2018
Language Technology for Language Documentation and Revitalization Workshop
> Designed software to assist language teachers searching small, unannotated corpora for specific linguistic phenomena. Publication in *STLU-CCRL*

Automated Transcription of Low-Resource Languages   summer 2018
Language Software Development (LSDev), SIL International
> Goal: build software to assist transcription of languages that have no established orthography. I initiated research on integrating speech recognition and on principles of interface design for users with low computer literacy. The software is under development.

Supersense Disambiguation of Prepositions   2018
Computational Language and Education Research (CLEAR)
> Focus: data-driven semantic disambiguation of prepositions and possessives. I co-developed the Semantic Network of Adposition and Case Supersense (SNACS) hierarchy and led the English annotation team to create a gold-standard corpus. Publication in *ACL*.

VerbNet for Constructing Worlds Computers (CWC)   2017
Computational Language and Education Research (CLEAR)
> Goal: to improve VerbNet's interoperability and syntactic accuracy. I redesigned its syntax representation to direct mapping between VerbNet and Universal Dependencies (UD).

Language Documentation and Community Development   2013-2015
In cooperation with Ufuq-S and SIL International
> I conducted documentary fieldwork on two threatened languages spoken in Azerbaijan, curating endangered data and surveying endangered linguistic and cultural practices. Two endangered corpora were preserved in long-term digital repositories.

Metaphor Project   2011-2013
Language Computer Corporation
> Aim: to automatically detect metaphors in Russian, Spanish, Farsi, and English. I developed annotation guidelines for defining and analyzing metaphors, mapped Russian dependency grammar to Stanford dependencies, and employed linguistic knowledge to improve Russian and Farsi syntactic parsers. These contributions led to the open access "Corpus of Rich Metaphor Annotation" at ISI USC and a co-authored publication in *CoNLL*.

TEACHING EXPERIENCE

Graduate Part-time Instructor   2020
Department of Linguistics, University of Colorado Boulder

Teaching Assistant   2016-2019
Department of Linguistics, University of Colorado Boulder

Adjunct Instructor   2011-2016
Dallas International University

Teaching Assistant   2010

Dallas International University

English as a Foreign Language Teacher / Russian Tutor   2002-2008
free-lance

ADVISING EXPERIENCE

Facilitator   2020-present
Graduate Research Workshop, CU Boulder
    Guiding junior Linguistics graduate students towards their professional development goals.

Supervisor   2018-present
Computational Language and Education Research (CLEAR)
    Mentoring 2-6 new research assistants each semester and leading annotation (semantic role labeling, named entity recognition, event primitives) teams in English, Russian, Ukrainian, Arabic, Arapaho, Spanish for DARPA-funded (KAIROS, AIDA, CWC, LORELEI, Metaphor) and NSF-funded (UMR) projects.

Peer Mentor   2017-present
University of Colorado Boulder
    Mentoring each year 1-2 first-year MA/PhD students in Linguistics and Computational Linguistics who are interested in pursuing a PhD.

Principal Investigator   2019-2020
Dissertation Research
    Mentored two undergraduate and one MA research assistant hired under a CARTSS grant. I trained the computer science undergraduates in basic linguistic analysis and guided the linguistics MA student in Python programming. Two students became co-authors on academic publications (cf. *EMNLP* publication) and one has continued related research on Arapaho.

Project Lead/Linguist   2011-2013
Language Computer Corporation
    Trained and superviseed native speaker annotators in linguistics and metaphor theory, developed annotation guidelines, and managed four annotation teams.

INDUSTRY POSITIONS

Project Manager   2018
LSDev, SIL International

Linguist   2011-2015
Language Computer Corporation

Student Intern   2009-2010
International Linguistics Office, SIL International

SERVICE

Secretary   2020-present
ACL Special Interest Group on Endangered Languages (SIGEL)

Program Chair   2020-2021
Fourth Workshop on Computational Methods for Endangered Languages (ComputEL-4)

Student Member   2019
Computational Linguistics Track Curriculum Development Committee, CU Boulder

Editorial Board   2018-present
Colorado Research in Linguistics (CRiL)

Ad hoc reviewer
North American Chapter of the Association for Computational Linguistics (NAACL)   *2021*
European Chapter of the Association for Computational Linguistics (EACL)   *2021*
Empirical Methods in Natural Language Processing (EMNLP)   *2020*
Language Resources and Evaluation Conference (LREC)   *2019, 2020*
Language Documentation & Conservation (LD&C)   *2019*

Program Committee   2019, 2020
International Workshop on Designing Meaning Representations (DMR)

Elected Student Representative   2019
Computational Linguistics Search Committee (for Assistant Professor), Dept. of Linguistics CU Boulder

Student Representative   2018
Dept. of Linguistics Faculty Board, CU Boulder

Chairman   2010
Student Body Association, Dallas Intl. University

Secretary   2009-2010
Student Body Association, Dallas Intl. University

Student Representative   2010
Board of Trustees, Dallas Intl. University

English as Second Language Teacher   multiple
Community classes for refugees and immigrants
St. Louis, MO 2000-2001; Dallas, TX 2010-2013

Service Learning Leader   2010
Dallas International University and Mississippi Band of Choctaw Indians

PROFESSIONAL MEMBERSHIPS

Association for Computational Linguistics (ACL)
ACL Special Interest Group for Endangered Languages

ACL Special Interest Group for Typology

FELLOWSHIPS, GRANTS, AND AWARDS

Research Grant, Institute of Cognitive Science, 2020

Individual Research Grant, International Literacy and Development (ILAD), 2020

Graduate Student Grant with supplemental award, Center for Advancement of Teaching in Social Sciences (CARTSS), 2019

Summer Research Fellowship, Dept. of Linguistics, CU Boulder, 2017

Summer Research Fellowship, Dept. of Linguistics, CU Boulder, 2018

Department Fellowship, Dept. of Linguistics, CU Boulder, 2016-2017

REFEREED PUBLICATIONS

Sarah Moeller, Ling Liu, and Mans Hulden. Under review. To POS Tag or Not to POS Tag: The Effect of POS Tags on Morphological Analysis in Low-Resource Settings. In *NAACL* 2021.

Jens E. L. Van Gysel, ... **Sarah Moeller**, ... Nianwan Xue. Under review. Designing a Uniform Meaning Representation for Natural Language Processing. *KI - Künstliche Intelligenz, Special Issue: NLP and Semantics.*

Sarah Moeller and Mans Hulden. 2021. Integrating Automated Segmentation and Glossing into Language Documentation. In *ComputEL-4*.

Moeller, Sarah, Ling Liu, Changbing Yang, Katharina Kann, and Mans Hulden. IGT2P: From Interlinear Glossed Texts to Paradigms. In *EMNLP*, pages 5251–5262, 2020a.

Graham Neubig, ... **Moeller, Sarah** ..., and Patrick Littell. A Summary of the First Workshop on Language Technology for Language Documentation and Revitalization. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 342–351, Marseille, France, 2020.

Moeller, Sarah, Irina Wagner, Martha Palmer, Kathryn Conger, and Skatje Myers. The Russian PropBank. In *LREC*, pages 5997–6004, Marseille, France, 2020b.

Kevin Stowe, **Moeller, Sarah**, Laura Michaelis, and Martha Palmer. Linguistic Analysis Improves Neural Metaphor Detection. In *CoNLL*, pages 362–371, Hong Kong, China, 2019.

Moeller, Sarah R. Customizing the LinGO Grammar Matrix Morphology. *Colorado Research in Linguistics (CRiL)*, 24, 2019.

Moeller, Sarah, Ghazaleh Kazeminejad, Andrew Cowell, and Mans Hulden. Improving Low-Resource Morphological Learning with Intermediate Forms from Finite State Transducers. In *Proceedings of the Workshop on Computational Methods for Endangered Languages (ComputEL-3)*, volume 1, Honolulu, HI, 2019.

Ling Liu, Ilamvazhuthy Subbiah, Adam Wiemerslage, Jonathan Lilley, and **Moeller, Sarah**. Morphological Reinflection in Context: CU Boulder's Submission to CoNLL-SIGMORPHON 2018 Shared Task. In *Proceedings of the CoNLL SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection*, pages 86–92, Brussels, 2018.

Moeller, Sarah, Ghazaleh Kazeminejad, Andrew Cowell, and Mans Hulden. A Neural Morphological Analyzer for Arapaho Verbs Learned from a Finite State Transducer. In *Proceedings of the Workshop on Computational Modeling of Polysynthetic Languages*, pages 12–20, Santa Fe, New Mexico, USA, 2018.

Moeller, Sarah and Mans Hulden. Automatic Glossing in a Low-Resource Setting for Language Documentation. In *Proceedings of the Workshop on Computational Modeling of Polysynthetic Languages*, pages 84–93, Santa Fe, New Mexico, USA, 2018.

Nathan Schneider, Jena D. Hwang, ... **Moeller, Sarah R.** ..., and Omri Abend. Comprehensive Supersense Disambiguation of English Prepositions and Possessives. In *ACL*, pages 185–196, Melbourne, Australia, 2018.

Moeller, Sarah Ruth. Word Stress in Lezgian in Optimality Theory. *Occasional Papers in Applied Linguistics (OPAL)*, 10:1–9, 2011.

Mekoulnodji Ndjerareou, Christy Melick, and **Moeller, Sarah**. A Brief Grammatical Sketch of Ngambay. *GIALens: Electronic Notes Series*, 4(2), 2011.

TEXTBOOK

Brenda Boerger, **Sarah Moeller**, D. Will Reiman, and Stephen Self. *Language and culture documentation manual*. LeanPub.com, 2016

TECHNICAL REVIEW

Sarah Moeller. Review of SayMore, a tool for language documentation productivity. *Language Documentation & Conservation*, 8:66-74, 2014.

PRESENTATIONS/INVITED TALKS

"Applications of Interlinear Glossed Texts" Guest lecture in field methods course at University of North Texas, 2020.

"SayMore X" Tech Showcase at 6th International Conference on Language Documentation and Conservation, University of Hawaii at Manoa, February 28-March 3, 2019.

Invited to present on digital tools at workshop on specifications for a metadata entry and collection management software, Alcanena, Portugal. October 2017.

"Developments in SayMore: The language documentation tool for citizen scientists" Paper presented at the 4th International Conference on Language Documentation and Conservation, University of Hawaii at Manoa, February 27-March 1, 2015.

"Ups and Downs: Bringing Students and Language Communities Together" Poster presented at the 2nd International Conference on Language Documentation and Conservation, University of Hawaii at Manoa, February 11-13, 2011, and 3rd Dallas-Ft. Worth Metroplex Linguistics Conference, University of North Texas (UNT), September 10, 2011.

"A preliminary analysis of Crimean Tatar written discourse." Paper presented at 2nd Dallas-Ft. Worth

Metroplex Linguistics Conference, Graduate Institute of Applied Linguistics, November 5, 2010.

CLASSES/WORKSHOPS TAUGHT

Programming for Linguistics	Instructor, University of Colorado	2020
Database Design, Annotating, and Archiving	Instructor, CoLang Workshop, U. of Alaska	2016
SayMore	Instructor, Transarch Workshop, Free University of Bolzano-Bozen	2015
Introduction to Grammatical Analysis	Instructor, Dallas Intl. University	2015
Language Documentation	Associate Instructor, Dallas Intl. University	2011-2015
Programming for Linguistics	Teaching Assistant, CU Boulder	2019
Languages of the World	Teaching Assistant, CU Boulder	2017-2019
Introduction to Linguistics	Teaching Assistant, CU Boulder	2016
Introduction to Grammatical Analysis	Teaching Assistant, Dallas Intl. University	2010
Second Language and Culture Acquisition	Teaching Assistant, Dallas Intl. University	2010

SKILLS

Languages	English (mother tongue)
	Russian (fluent)
	German (conversant)
	Georgian, French, Latin, Koine Greek (beginner)
	Lezgi, Ngambay, Tsakhur, Arapaho (structural knowledge)
Technical	Python (scikit-learn, Keras), regex, LaTeX, Figma. Familiar with: R, Java, C, SQL