David R. Mortensen

1029 LaClair Street, Pittsburgh PA 15218-1226

+1 412 527 8827 | dmortens@cs.cmu.edu | http://www.davidmortensen.org

---

Executive Summary

A linguist who raises performance. Known for defying disciplinary boundaries in pursuit of optimal solutions to natural language problems. Blends academic expertise in the language sciences and background in diverse methods of data analysis with formidable scripting and database skills.

Professional Experience

Carnegie Mellon University; Pittsburgh, PA
Research Scientist, Aug 2015–present
Applied linguistic knowledge, data annotation expertise, and data management skills to the development of NLP systems.

- Provided linguistic support to a team participating in DARPA's LORELEI program, focused on developing low resource language human language technologies for humanitarian assistance/disaster relief scenarios (HA/DR).
- Developed an ontology of HA/DR scenarios and associated keywords.
- Supervised corpus annotation for named entity recognition and a novel information extraction task (situation frame extraction).
- Developed corpus annotation training materials.
- Participated in the development of a state-of-the-art phonologically-aware neural-CRF named entity recognition system.
- Developed massively multilingual orthography-to-IPA conversion component (Epitran).
- Developed IPA-articulatory feature database and accompanying Python library (PanPhon).
- Supervised error analysis of NER system output in Mandarin Chinese, Uyghur, Tigrinya, and Oromo.
- Developed phonological component of typological database designed for use in NLP tasks (URIEL).
- Taught courses on NLP and computational linguistics.

Carnegie Mellon University; Pittsburgh, PA
Visiting Researcher, May 2014-Sep 2014
Engaged with two teams in research on computational linguistics and NLP.

- Refined rules for converting a treebank of constituency trees to dependency trees.
- Collaborated with computer scientists in the rapid development of named entity recognition systems for under-resourced languages.
- Developed FST for conversion of lexemes between two closely related minority languages.
- Participated in the training and evaluation of a weakly supervised morphological analyzer.

University of Pittsburgh; Pittsburgh, PA
Assistant Professor, 2006–2014

Carried out a research and teaching program in linguistics.

- Published linguistics research in peer-reviewed journals.
- Carried out linguistic fieldwork on underdescribed languages.
- Performed corpus-based, quantitative and qualitative research on word structure in a minority language.
- Taught four courses per year in various subfields of linguistics, including phonetics, phonology, morphology, field methods, and linguistic typology.
- Trained undergraduate and graduate students in theoretical and field linguistics, cultivating communication and mentoring skills.
- Supervised research assistants, developing management skills.
- Supervised two MA thesis, one PhD dissertation, and participated in numerous MA and PhD committees.

University of California, Berkeley; Berkeley, CA
Graduate Student Instructor, 2004–2005
Assisted in the instruction of large undergraduate classes in linguistics.

- Developed excellent pedagogical skills.
- Earned *Outstanding Graduate Student Instructor Award.*

University of California, Berkeley; Berkeley, CA
Graduate Student Researcher, 2003–2006
Developed computational tools to maintain, manipulate, and query a set of etymological and lexical databases.

- Developed, as part of a team, a web application (Perl/MySQL) for etymological analysis and management of comparative linguistic data.
- Migrated a large, multilingual lexical database from FileMaker to MySQL.
- Authored tools for organizing and typesetting an English-Lahu dictionary (Python/LaTeX).

Skills

Languages: English (native), Hmong (advanced), Mandarin Chinese (intermediate), French (reading knowledge), Tangkhul (novice), German (novice)
Programming languages: Python, JavaScript, Haskell, Perl, R, Ruby, Java
Operating systems: Linux, MacOS, Windows
Other software: XFST, XLE, NLTK, Praat, Git, Bash, LaTeX, MySQL, PostgreSQL, SQLite
Other technologies: Unicode, SQL, XML/HTML, CSS

Education

- PhD in Linguistics. 2006. University of California, Berkeley. Dissertation: *Logical and Substantive Scales in Phonology.*
- MA in Linguistics. 2003. University of California, Berkeley.
- BA in English Literature (Departmental Honors, minors in Linguistics and Psychology). 2000. Utah State University.