DAVID L. KING

(+1)502.509.7360 ◇ david@dlk.ai ◇ https://dlk.ai/

SUMMARY

Computational linguist with 6+ years of industry experience in machine translation, knowledge graphs, and generation. Experience leading interdisciplinary teams across academia, industry, and multiple continents. Mentored over 20 junior colleagues.

EDUCATION

The Ohio State University, Columbus, OH *August 2014 - Present*
Doctorate of Philosophy, Computational Linguistics (expected August 2024)
Master of Arts, Computational Linguistics (2018)
Department of Linguistics

University of Kentucky, Lexington, KY *August 2004 - May 2008*
Bachelor of Arts, Linguistics and German
Minors in Arabic and Sanskrit

EXPERIENCE

Vantage Labs, LLC, New Hope, Pennsylvania *January 2024 - Present*
Computational Linguist: Converted rule-based ESL grammar checker to PyTorch using sequence classification models. Generated data for sparse error classes using Llama2/3. Optimized models for CPU inference using ONNX and OpenVINO. Currently building new team.

NWO.ai, New York, New York *December 2021 - October 2023*
Lead NLP Research Scientist: Prototyped and implemented knowledge graph used for data mining micro-trends in social media, patents, newswire, and scientific data. Implemented Chinese MT pipeline for processing Weibo data. Designed entity linking for processing 4B social media posts. Prototyped ontology induction model with graph embedding. Built ML/NLP team.

Lengoo, GmbH, Berlin, Germany *May 2018 - September 2019*
Senior Computational Linguist: Implemented sequence-to-sequence neural machine translation framework and pipeline. Transitioned company from a traditional translation service provider to using machine translation, cutting translation turnaround by half. Facilitated hiring decisions for new machine learning department.

The Ohio State University, Columbus, Ohio *August 2015 - December 2021*
GRA: Research assistant for John Glenn College of Public Affairs: Developed data gathering and transformer based classifier for detecting state-sponsored COVID-19 misinformation in social networks.
GRA: Research assistant for Michael White. Projects included Madly Ambiguous (http://madlyambiguous.osu.edu/), recovering long-distance dependencies (CEUDO), and building a virtual standardized patient question-answering system for training first-year medical students.
GTA: Teaching assistant for Linguistics 3802: Language and Computers.

Appriss, Inc., Louisville, Kentucky *October 2012 - August 2014*
UX Designer and Field Service Engineer: Developed HEALTH MONITORING TOOL, a data mining tool for monitoring code deployment and predicting victim notification service (VINE, SAVN, SAVAN) outages across 48 states and Puerto Rico. Also redesigned and updated Mandarin interface for California.

Peace Corps, Lanzhou, Gansu, P. R. China *July 2010 - July 2012*
Volunteer: Primary instructor for ESL (1000+ hours), grant writing (RELO), and Linux computer lab founder and administrator. Also founded the ENGLISH RESOURCE CENTER, a physical library of 200+ books and English learning resources.

SKILLS

Programming, Machine Learning, and Related Frameworks
Python, Bash, and PyTorch—Primary
R, Ruby, Prolog, Neo4J, Torch (Lua), Theano, and Tensorflow
HPC and Cloud Platforms
SLURM, Moab, AWS, and Google Cloud
Common NLP Toolkits/Packages
BERT, ELMo, Word2Vec, GloVe, WordNet, PPDB, SpaCy, NLTK, Sklearn, SciPy, ONNX, OpenVINO, and Triton
Primary Languages
English—Native language
German—Working proficiency
Mandarin—ACTFL advanced-mid rating

GRANTS AND AWARDS

IARPA COVID-19 Research Topic Grant

Battelle Seedling Project Grant

Ohio Supercomputer Flash Talk Funding

Targeted Investment in Excellence Grant

Invited Speaker Funding from RASA GmbH

Techstar Startup Funding for Lengoo GmbH

OPEN SOURCE SOFTWARE (SELECTED LIST)

SCiL-20 (`https://github.com/DavidLKing/SCiL-20`)
A series of morphological analysis tools I wrote to investigate morphological sequence-to-sequence models and create more linguistically salient (i.e. explainable) analyses for the errors they produce.

pytorch-MED (`https://github.com/DavidLKing/MED-pytorch`)
My reimplemention of Kann and Schūtze's 2016 Morphological Encoder Decoder written in PyTorch and IBM's `pytorch-seq2seq` framework.

correctMalt (`https://github.com/DavidLKing/correctMalt`)
Script for correcting the SIGMORPHON 2016 Maltese data for the Unimorph Consortium. Designed to make the 2016 SIGMORPHON data reflect the same annotation in the Gabra database.

CEUDO (`https://github.com/DavidLKing/CEUDO`)
A platform to integrate the Stanford Dependency Converter (SDC) output with CCGbank to get a representation closer to Universal Dependencies as described in Nivre et al. 2016 and in the manual at universaldependencies.org.

PEER REVIEWED PROCEEDINGS

David L. King, Andrea D. Sims, Micha Elsner. Interpreting Sequence-to-Sequence Models for Russian Inflectional Morphology. 2020. In Proc. of the Society of Computation in Linguistics at LSA 2020.

Micha Elsner, Andrea D. Sims, Alexander Erdmann, Antonio Hernandez, Evan Jaffe, Lifeng Jin, Martha Booker Johnson, Shuan Karim, David L. King, Luana Lamberti Nunes, Byung-Doh Oh, Nathan Rasmussen, Cory Shain, Stephanie Antetomaso, Noah Diewald, Kendra V. Dickinson, Michelle McKenzie, and Symon Stevens-Guille. Modeling Morphological Learning, Typology, and Change: What can the neural sequence-to-sequence framework contribute? Journal of Language Modelling 7.1 (2019): 53-98.

Kartikeya Upasani, David L. King, Jinfeng Rao, Anusha Balakrishnan, and Michael White. 2019. The OSU-Facebook Realizer for SR '19: Seq2seq Inflection and Serialized Tree2Tree Linearization. In Proc. of the Workshop on Multilingual Surface Realization at EMNLP-IJNLP 2019.

David L. King and Michael White. 2018. The OSU Realizer for SRST 18: Neural Sequence-to-Sequence Inflection and Incremental Locality-Based Linearization. In Proc. of the Workshop on Multilingual Surface Realization at ACL 2018.

Taylor Mahler, Willy Cheung, Micha Elsner, David L. King, Marie-Catherine de Marneffe, Cory Shain, Symon Stevens-Guille, and Michael White. Breaking NLP: Using Morphosyntax, Semantics, Pragmatics and World Knowledge to Fool Sentiment Analysis Systems. In Proc. at the Build it. Break it. NLP Workshop at EMNLP 2017

Michael White, Manjuan Duan, and David L. King. A Simple Method for Clarifying Sentences with Coordination Ambiguities. In Proc. at Explainable Computational Intelligence Workshop at INLG 2017

David L. King and Michael White. Enhancing PTB Universal Dependencies for Grammar-Based Surface Realization. In Proc. at INLG 2016.

David L. King. Evaluating Sequence Alignment for Inflectional Morphology. In Proc. at the Special Interest Group for Phonology and Morphology Workshop at ACL 2016.

Evan Jaffe, Lifeng Jin, David L. King, and Marten van Schijndel. Azmat: Sentence Similarity using Associative Matrices. In Proc. at the International Workshop on Semantic Evaluation at NAACL 2015.

ACTIVITIES

Invited Talks

Surrey Morphology Circle, University of Surrey, UK *TBA*
Learning Syncretism: What Computational Models Can Tell us about Typological Claims

The 4th Annual International Chatbot Summit, Berlin *2018*
Neural Sequence-to-Sequence Inflection and Incremental Locality-Based Linearization

Machine-Learning Learning Group, Berlin *2018*
Neural Sequence-to-Sequence Inflection and Incremental Locality-Based Linearization

Scholarly Presentations

David L. King, Andrea Sims, and Micha Elsner. October 17, 2019. Sequence-to-Sequence Learning for Russian Inflectional Morphology. Ohio Supercomputer User Group Conference, Columbus, OH.

Sarah Ewing, Amad Hussain, David L. King, and Michael White. May 3, 2019. Ranking Automatic Paraphrases with Contextualized Word Embeddings. Midwest Speech and Language Days 2019, Chicago.

David L. King, Andrea Sims, and Micha Elsner. September 14, 2018. Sequence-to-Sequence Learning for Russian Inflectional Morphology. Center for Cognitive and Brain Sciences, Mt. Sterling, OH.

David L. King, Andrea Sims, and Micha Elsner. May 4, 2017. Sequence-to-Sequence Learning for Russian Inflectional Morphology. Midwest Speech and Language Days 2017, Chicago.

David L. King and Michael White. May 13, 2016. Improving Universal Dependency Output of the Penn Treebank. Midwest Speech and Language Days 2016, Bloomington, IN.

SERVICE

Peer Review

IJCNLP *2023*
No specific track

IJCNLP *2022*
No specific track

EACL *2021*
Track: Summarization and Generation

EMNLP-IJNLP *2020*
Track: Summarization and Generation

EMNLP-IJCNLP *2019*
Track: Summarization and Generation

ACL *2019*
Track: Machine Learning for NLP

Tutorials
"Introduction to Using Unix" *2014, 2019*
"Introduction to Tensorflow" *2016*

Committees
American International Morphology Meeting (Program Committee) *2021*
Laboratories and Computing *2014-2020*
Travel *2017-2020*
Diversity *2015-2016*
Speakers *2014-2016*

Guest Lectures
Language and Computers, invited by Evan Jaffe *2019*
Supplementing Sequence to Sequence Learning for Inflectional Morphology

Language and Computers, invited by Evan Jaffe *2019*
Neural Sequence-to-Sequence Inflection and Incremental Locality-Based Linearization

MENTORING AND TRAINING

Undergraduate
Ronnie Eytchison *May 2020 - December 2021*
John Kostik *May 2020 - December 2021*
Cheng Zhang *August 2018 - May 2019*
Sarah Ewing *January 2018 - May 2018*
Amad Hussain *August 2017 - December 2018*

Graduate
Rod Abhari *November 2020 - December 2021*
Ashley Lewis *October 2019 - Present*

Post-Doctoral
Yunkang Yang *August 2020 - November 2020*
Matthew Osborne *August 2020*
Xintong Li *December 2019*

Industry
Mert Alev *February 2022-March 2023*
Culton Koster *April 2022-Present*
Tina Tiwari *January 2023-August 2023*
Ivona Kocheva *July 2023-October 2023*