

Sijia Ge

📧 konic-nlp.github.io @ sijiage007@gmail.com 🎓 [Google Scholar](#) in linkedin.com/in/sijia-ge

Education

University of Cincinnati , Cincinnati, OH PhD in Computer Science > <u>Research Interests</u> : Computational Linguistics, Alignment, Multilingual NLP, Mechanistic Interpretability	08/2025 – present
University of Colorado Boulder , Boulder, CO M.S. in Computational Linguistics	08/2021 – 12/2023
Nanjing Normal University , China M.A. in Applied Linguistics (Language Processing Track)	09/2016 – 06/2019
Shanxi University , China B.A. in Chinese Language and Literature	09/2012 – 07/2016

Experience

Riphey.AI , Colorado, USA ML/AI Engineer	01/2024 – 02/2025
University of Colorado , Colorado, USA Student Researcher with Martha Palmer	05/2022 – 12/2023
Beijing Lingosail Technology Co., Ltd. , China Product Manager Intern	03/2021 – 06/2021
Nanjing Normal University , China Student Researcher with Bin Li	07/2017 – 01/2019

Manuscript under review

- [1] **Is a Flying Penguin a Typical Bird? Probing Prototype Judgments in Large Language Models**
Sijia Ge et al. [Manuscript under review]

- [2] **Do LLMs Trade Information for Cost? A Cost-Aware Guessing Game**
Sijia Ge et al. [Manuscript under review]

Publications

- [1] **Annotate Chinese Aspect with UMR—a Case Study on the Little Prince** 📄
[Sijia Ge](#), Zilong Li, Alvin Po Chun Chen, Guanchao Wang
Language Resources and Evaluation [LREC 2024]

- [2] **Building a Broad Infrastructure for Uniform Meaning Representations** 📄
Julia Bonn, Matthew J Buchholz, Jayeol Chun, Andrew Cowell, William Croft, Lukas Denk, [Sijia Ge](#), Jan Hajic, Kenneth Lai, James H Martin, Skatje Myers, Alexis Palmer, Martha Palmer, Claire Benet Post, James Pustejovsky, Kristine Stenzel, Haibo Sun, Zdenka Urešová, Rosa Vallejos, Jens EL Van Gysel, Meagan Vigus, Nianwen Xue, Jin Zhao
Language Resources and Evaluation [LREC 2024]

- [3] **PropBank goes Public: Incorporation into Wikidata** 📄
Elizabeth Spaulding, Kathryn Conger, Anatole Gershman, Mahir Morshed, Susan Windisch Brown, James Pustejovsky, Rosario Uceda-Sosa, [Sijia Ge](#), Martha Palmer
Linguistic Annotation Workshop [LAW XVIII]

- [4] **UMR-Writer 2.0: Incorporating a new keyboard interface and workflow into UMR-Writer** 📄
[Sijia Ge](#), Jin Zhao, Kristin Wright-Bettner, Skatje Myers, Nianwen Xue, Martha Palmer
Linguistic Annotation Workshop [LAW XVII]

- [5] **Integration of Named Entity Recognition and Sentence Segmentation on Ancient Chinese based on Siku-BERT** 
 Sijia Ge
International Workshop on Natural Language Processing for Digital Humanities [NLP4DH 2022]
- [6] **An easier and efficient framework to annotate semantic roles: Evidence from the Chinese AMR corpus** 
 Li Song, Yuan Wen, Sijia Ge, Bin Li, Weiguang Qu
Linguistic Annotation Workshop [LAW XIII]
- [7] **The Construction and Analysis of Annotated Imagery Corpus of Three Hundred Tang Poems** 
 Xingyue Hao, Sijia Ge, Yang Zhang, Yuling Dai, Peiyi Yan, Bin Li
Chinese Lexical Semantics Workshop [CLSW 2019]

Featured Projects

AI-assisted Training Application Prototype

05/2026–06/2026

University of Cincinnati, in collaboration with Procter & Gamble

- › Developed a full-stack employee-training prototype using React and Flask.
- › Integrated an LLM with a retrieval-augmented generation (RAG) pipeline to answer training-related questions.
- › Supported multimodal instructional content and embedding-based query clustering to improve information access and knowledge organization.

Email & Document Automation in Logistics

01/2024 – 02/2025

Rippey.AI

- › Applied template and overlapping pixel matching to semi-automate document data annotation.
- › Transformed email extraction tasks into Q&A tasks, effectively grouping weights and dimensions for individual items.

Uniform Meaning Representation Aspect Annotation on Chinese

08/2023 – 03/2024

University of Colorado Boulder

- › The Uniform Meaning Representation (UMR) adopts a lattice structure to capture various aspects of events, combining traditional *grammatical aspects* with *situation types*.
- › Evaluated the aspect lattice by annotating Chinese, conducted the first comparative analysis of UMR aspect annotation, grammatical aspect annotation, and situation-type annotation performed independently. [LREC'24]

Uniform Meaning Representation Annotation Tool

11/2022 – 12/2023

w/ Jin Zhao, Kristin Wright-Bettner, Skatje Myers, Nianwen Xue, Martha Palmer

University of Colorado Boulder CLEAR Lab

- › Developed an innovative keyboard-based interface for Semantic Role Labeling annotation, reducing effort by 60% compared to a click-based interface.
- › Supported multiple languages, including Chinese and Arabic. [LAW XVII]

Ancient Chinese Named Entity Recognition and Sentence Segmentation

03/2018 – 12/2018

w/ Chengming Li, Ning Cheng, Bin Li

Nanjing Normal University

- › Developed a unified framework that integrates sentence segmentation and named entity recognition into a single sequence labeling task.
- › Demonstrated that the joint framework outperformed separate models for the two tasks. [LT4HALA '20]

Imagery Annotation on Tang Poems

10/2018 – 04/2019

Nanjing Normal University

- › Annotated imagery includes both its literal meaning, represented by the surface form, and its implicit metaphorical meaning. Additionally, annotations capture the connections between these two meanings, such as similarities, relevance, and other relationships.
- › Developed a corpus encompassing over 300 Tang poems. [CLSW '19]

Teaching

CS 4092 Database Design and Development , University of Cincinnati. <i>Teaching Assistant</i>	2026 Spring
CS 5154&6064 Information Retrieval , University of Cincinnati. <i>Teaching Assistant</i>	2025 Fall
CSCI 5622 Machine Learning , University of Colorado Boulder. <i>Course Manager w/</i> David Quigley	2023 Spring
CSCI 4622 Machine Learning , University of Colorado Boulder. <i>Course Manager w/</i> Ravi Starzl	2023 Fall

Academic Service

Reviewer	NLP4DH & IWCLUL'23, NLP4DH'25, NLP4DH'26
Volunteer	ACL'23, DSNA '23, AACL'22, CCL'17