

Chenyi (Skylar) Ding

cyiding@ucdavis.edu

Research Interests: computational psycholinguistics, surprisal theory, L2 sentence processing, word learning, language acquisition

EDUCATION

University of California, Davis

Master of Arts in Linguistics

Davis, CA

Sep 2025 – Expected Jun 2027

- GPA: 3.96/4.00; Advisor: Emily Morgan

Xi'an JiaoTong-Liverpool University (XJTLU)

Bachelor of Arts in English Studies

Suzhou, China

Sep 2021 – Jul 2025

- Upper Second Class Honours (2:1)
- Thesis: “The Joint Influence of Proficiency and Cognitive Processing Efficiency on L2 Predictive Sentence Processing” (Supervisor: Hong Liu)

PUBLICATIONS

Peer-reviewed

Ding, C., Houghton, Z. N., & Jasbi, M. (2026). Reading acquisition order out of BabyLM representations: Child-scale training produces developmentally aligned semantic spaces. *Proceedings of the BabyLM Workshop at EMNLP 2026*, Budapest, Hungary. Forthcoming.

Under review

Ding, C., & Liu, H. The joint influence of proficiency and cognitive processing efficiency on L2 predictive sentence processing. *Second Language Research*.

PRESENTATIONS

Ding, C., & Morgan, E. Decomposing language model surprisal reveals differential cue weighting in L1 and L2 reading. 2027 Annual Meeting of the Linguistic Society of America (LSA), San Francisco, CA, Jan 2027. Oral presentation.

Ding, C., Houghton, Z. N., & Jasbi, M. Children’s word-learning order is already in their input: Age of acquisition in the semantic space of a child-scale language model. **51st Boston University Conference on Language Development (BUCLD 51)**, Boston, MA, Nov 2026. Poster.

Ding, C., Chen, S., Zhao, Z., Ye, M., & Frey, S. Measuring pragmatic difficulty in LLM-based content analysis: Semantic entropy as a diagnostic for variable-level reliability. **12th International Conference on Computational Social Science (IC²S²)**, Burlington, VT, Jul 2026. Talk.

Ding, C., Chen, S., Zhao, Z., Ye, M., & Frey, S. Why here and not there? A pragmatic account of selective failures in LLM-based construct extraction. **Communication Horizons @ UC Davis 2026: Communication in the Age of AI and Algorithms**, Davis, CA, May 2026. Talk.

Ding, C., Chen, S., Zhao, Z., Ye, M., & Frey, S. Pragmatic vulnerability in academic discourse: Using semantic entropy to predict where LLMs fail. **12th Annual UC Davis Symposium on Language Research (CLR)**, Davis, CA, May 2026. Talk.

SKILLS

Programming: Python, R

Statistics: Linear/logistic regression, mixed-effects models, Bayesian methods

Experimental: Eye-tracking (Tobii Pro Lab, Tobii Pro Spectrum, EyeLink), Gorilla, PsychoPy, E-Prime, Praat, Audacity, Qualtrics

Languages: Chinese (native), English (proficient), Japanese (basic)

RESEARCH EXPERIENCE

Age of Acquisition in Child-Scale Language Models

Mar 2026 – present

Lead Researcher, Department of Linguistics, UC Davis (Advisor: Masoud Jasbi; with Zachary Houghton)

- Designed and led a study relating children’s word-learning order to the semantic structure of child-directed input, using language models trained on child-scale corpora (BabyLM) as models of the input.
- Compared 18 public checkpoints across three architecture families (OPT, GPT-2, RoBERTa) and data scales from 10M to 180B words; extracted contextual embeddings over CHILDES utterances (PyTorch, Hugging Face) and modeled Wordbank age-of-acquisition norms for 600+ CDI words with cross-validated ridge regression.

- Developed three measures of semantic-neighborhood structure (cohesion, distinctiveness, diversity) from embedding geometry and tested them against the classic predictors with hierarchical regression and Bayesian models (brms).
- Built a corpus-based age-of-acquisition measure from children's first productions in CHILDES (R, childesr). First-authored the resulting paper, accepted at the BabyLM Workshop (EMNLP 2026), and a poster accepted at BUCLD 51.

Cue-Weighted Surprisal: Decomposing Language Model Surprisal in L1 and L2 Reading Apr 2026 – present
Graduate Student Researcher (Summer 2026), Department of Linguistics, UC Davis (Advisor: Emily Morgan)

- Designed a study testing the Shallow Structure Hypothesis with language model surprisal, decomposing GPT-2 surprisal into syntactic and semantic components via the chain rule, with word frequency as a separate predictor.
- Computed syntactic surprisal by summing GPT-2 predicted probabilities over words sharing a part of speech, and built unigram, n-gram, and GloVe-similarity baselines for comparison.
- Modeled first-pass reading times from the Multilingual Eye-movement Corpus (45 L1 and 497 L2 readers from 12 language backgrounds) with linear mixed-effects models including by-subject random slopes and by-item intercepts.
- Extended the analysis to per-L1-background comparisons across 10 language groups. First-authored abstracts submitted to LSA 2027 and CAMP 9.

Pragmatic Vulnerability in LLM-Based Content Analysis Jan 2026 – present
Researcher, Department of Communication, UC Davis (Advisor: Seth Frey; with Shiyi Chen, Ziyu Zhao, and Miner Ye)

- Developed a four-dimensional framework drawing on common ground theory, Gricean pragmatics, and register analysis to classify content-analysis variables as pragmatically vulnerable or routine.
- Applied a GPT-4o-mini extraction pipeline to 55 peer-reviewed articles on three audience-engagement constructs, extracting 57 codebook variables across five independent runs to measure output consistency.
- Adapted semantic entropy as a variable-level diagnostic: sampled ten extractions per article–variable pair, clustered them by bidirectional entailment, and modeled entropy with crossed-random-effects mixed models. Presented the work at IC²S² 2026, Communication Horizons, and the UC Davis CLR Symposium.

Proficiency and Cognitive Efficiency in L2 Predictive Sentence Processing Sep 2024 – Jul 2025
B.A. Thesis Researcher, Department of Applied Linguistics, XJTLU (Supervisor: Hong Liu)

- Designed, with my supervisor, a visual world paradigm eye-tracking experiment on predictive processing in Chinese–English bilinguals, adding a flanker task to measure cognitive processing efficiency.
- Ran a norming study with 33 participants to validate the stimuli (semantic association between agent–verb combinations and referent images), then recruited and tested 143 participants in the eye-tracking experiment (Tobii Pro Spectrum, 1200 Hz).
- Analyzed fixation-based (log-gaze ratio) and saccade-based measures of anticipation with mixed-effects models in R. First-authored the manuscript (Ding & Liu), now under review at Second Language Research.

The Influence of Sentence Predictability on Lexical Processing among L2 Learners Mar 2024 – Jun 2025
Research Assistant, Department of Applied Linguistics, XJTLU (PIs: Darcy Sperlich and Hong Liu)

- Designed reading stimuli calibrated to B1 CEFR proficiency using corpus frequency data and ran a norming study (N = 60) to establish cloze predictability of target words.
- Recruited and screened participants with the Quick Placement Test, then administered a sentence-reading eye-tracking experiment to 80 participants in Tobii Pro Lab.
- Contribution acknowledged in Liu, Sperlich, & Yao (2026), *Linguistic Approaches to Bilingualism*, <https://doi.org/10.1075/lab.26019.liu>.

TEACHING EXPERIENCE

Department of Linguistics, University of California, Davis Fall 2025 – Fall 2026
Teaching Assistant, “Introduction to Linguistics” (Fall 2025, Winter 2026), “English Grammar” (Spring 2026), “Computational Linguistics” (Fall 2026); Grader, “Topics in Sign Language Linguistics” (Fall 2026)

- Led discussion sections, held office hours, and graded assignments and exams.

SERVICE AND MEMBERSHIPS

Reviewer: BabyLM Workshop 2026 (EMNLP); International Conference on Computational Social Science (IC²S²); UC Davis Symposium on Language Research (CLR)

Member: Linguistic Society of America (LSA)