Zhenghao "Herbert" Zhou

+1-314-327-8038 | herbert.zhou@yale.edu | herbert-zhou.github.io

New Haven, CT 06511, USA

Research Interests

- **Broad:** computational linguistics; psycholinguistics; cognitive science; artificial intelligence
- **Narrow:** mechanistic interpretability; sentence production and comprehension;

Education

- **Yale University** *August 2022 - Expected 2028*
  PhD Student in Linguistics New Haven, CT
  - Advisor: Robert Frank and R. Thomas McCoy (co-advising)
- **Washington University in St. Louis** *August 2018 - May 2022*
  BS in Computer Science & Mathematics St. Louis, MO
  - Second major in Philosophy-Neuroscience-Psychology (PNP Program); Minor in music
  - GPA: 3.98 / 4.00; *summa cum laude*
- **LSA Summer Institute** *Eugene, OR; July 2025*
- **European Summer School in Logic, Language and Information** *Leuven, Belgium; August 2024*
- **LSA Summer Institute** *Amherst, MA; June 2023*

Research Affiliations

- **Computational Linguistics at Yale (CLAY Lab)**, member and project leader *2022-present*
- **Yale DYNAMICS Lab**, member *2024-present*
- **Yale Language and Brain Lab**, member *2024*
- **WUSTL Computer Science and Engineering REU Program**, research assistant *2021*

Publications

[1] **Zhenghao Herbert Zhou** and Maria M. Piñango (2025). More Than One Type of Number Agreement Computation Process? Investigating Planning Time-Course through Agreement Attraction in Reflexive Pronoun Production. *Work in progress* 2025.

[2] **Zhenghao Herbert Zhou**, R. Thomas McCoy, and Robert Frank(2025). Causal Interventions on Continuous Features in LLMs: A Case Study in Verb Bias. In *First Workshop on CogInterp: Interpreting Cognition in Deep Learning Models (CogInterp)*, accepted, December 2025.

[3] Paul Smolensky, Roland Fernandez, **Zhenghao Herbert Zhou**, Mattia Opper, Adam Davies, and Jianfeng Gao (2025). Mechanism of Symbol Processing for In-Context Learning in Transformer Networks. In *Journal of Artificial Intelligence Research (JAIR)*, accepted, 2025. Preprint available at: https://arxiv.org/abs/2410.17498.

[4] Xiaomeng Zhu *, **Zhenghao Herbert Zhou** *, Simon Charlow, and Robert Frank (2025). Meaning Beyond Truth Conditions: Evaluating Discourse Level Understanding via Anaphora Accessibility. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 8824–8842, Vienna, Austria. Association for Computational Linguistics (ACL).*, July 2025.

[5] **Zhenghao Herbert Zhou**, Robert Frank, and R. Thomas McCoy (2025). Is In-Context Learning a Type of Error-Driven Learning? Evidence from the Inverse Frequency Effect in Structural Priming. In *In Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL).*, April 2025.

[6] **Zhenghao Herbert Zhou** and Robert Frank (2023). What affects Priming Strength? Simulating Structural Priming Effect with PIPS. In *Proceedings of the Society for Computation in Linguistics 2023 (SCiL).*, June 2023.

[7] Michael Wilson, **Zhenghao Herbert Zhou**, and Robert Frank (2023). Subject-verb agreement with Seq2Seq transformers: Bigger is better, but still not best. In *Proceedings of the Society for Computation in Linguistics 2023 (SCiL).*, June 2023.

Presentations

[1] Xiaomeng Zhu *, **Zhenghao Herbert Zhou** *, Simon Charlow, and Robert Frank (2025). Meaning Beyond Truth Conditions: Evaluating Discourse Level Understanding via Anaphora Accessibility. *Poster presented at the 63rd Annual Meeting of the Association for Computational Linguistics (ACL)*, Virtual Presentation, July 2025.

[2] Xiaomeng Zhu *, **Zhenghao Herbert Zhou** *, Simon Charlow, and Robert Frank (2025). Do LLMs Understand Anaphoric Accessibility?. *Poster presented at the Society for Computation in Linguistics 2025 (SCiL)*, Eugene, OR, Jul 2025

[3] **Zhenghao Herbert Zhou**, R. Thomas McCoy, and Robert Frank (2025). Compressing Structural Priming in Large Language Models through Function Vectors. *Poster presented at the Linguistics Society of America Summer Institute 2025*, Eugene, OR, Jul 2025

[4] **Zhenghao Herbert Zhou** and Maria M. Piñango (2025). More Than One Type of Number Agreement Computation Process? Investigating Planning Time-Course through Agreement Attraction in Reflexive Pronoun Production. *Poster presented at the Linguistics Society of America Summer Institute 2025*, Eugene, OR, Jul 2025

[5] R. Thomas McCoy and **Zhenghao Herbert Zhou** (2025). Understanding the abilities of AI systems. *Poster presented at the Envisioning AI at Yale: An Interdisciplinary Symposium 2025*, New Haven, CT, May 2025

[6] R. Thomas McCoy and **Zhenghao Herbert Zhou** (2025). How to evaluate large language models: Insights from linguistics. *Poster presented at the Envisioning AI at Yale: An Interdisciplinary Symposium 2025*, New Haven, CT, May 2025

[7] **Zhenghao Herbert Zhou**, Robert Frank, and R. Thomas McCoy (2025). Is In-Context Learning a Type of Error-Driven Learning? Evidence from the Inverse Frequency Effect in Structural Priming. *Talk presented at the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)*, Albuquerque, NM, Apr 2025

[8] **Zhenghao Herbert Zhou** (2025). Error-Driven Learning in DFT: A case study of structural priming. *Talk presented at the Linguistic Society of America 2025 Annual Meeting (LSA)*, Philadelphia, PA, Jan 2025

[9] **Zhenghao Herbert Zhou** (2024). Is In-Context Learning a Type of Error-Driven Learning Mechanism? Evidence from the Inverse Frequency Effects in Structural Priming. *Invited talk presented at the Hollymartin Lab at the University of Edinburgh*, Edinburgh, UK, Dec 2024

[10] **Zhenghao Herbert Zhou**, Robert Frank, and R. Thomas McCoy (2024). Is In-Context Learning a Type of Error-Driven Learning Mechanism? Diagnosing with the Inverse Frequency Effects in Structural Priming. *Poster presented at the 30th Architecture and Mechanisms for Language Processing (AMLaP)*, Edinburgh, UK, Sep 2024

[11] **Zhenghao Herbert Zhou**, Robert Frank, and R. Thomas McCoy (2024). Language Models Show Gradient Inverse Frequency Effects in Structural Priming: Implications for In-Context Learning. *Talk presented at the Student Session of the 35th European Summer School in Logic, Language and Information (ESSLLI)*, Leuven, Belgium, Aug 2024

[12] Michael Wilson, **Zhenghao Herbert Zhou**, and Robert Frank (2023). Subject-verb agreement with Seq2Seq transformers: Bigger is better, but still not best. *Poster presented at the Linguistics Society of America Summer Institute 2023, Amherst*, MA, Jul 2023

[13] **Zhenghao Herbert Zhou** and Robert Frank (2023). What affects Priming Strength? Simulating Structural Priming Effect with PIPS. *Talk presented at the Society for Computation in Linguistics 2023 (SCiL)*, Amherst, MA, Jun 2023

[14] Michael Wilson, **Zhenghao Herbert Zhou**, and Robert Frank (2023). Subject-verb agreement with Seq2Seq transformers: Bigger is better, but still not best. *Poster presented at the Society for Computation in Linguistics 2023 (SCiL)*, Amherst, MA, Jun 2023

Fellowships and Awards

- **Society of Computation in Linguistics Student Travel Award**, $500 *2025*
- **Yale Linguistics Department Conference Travel Grants**, $1000 per year *2022-present*
- **Yale MacMillan Center Summer Language Grant**, $2500 per year *2023, 2025*
- **Yale University Sterling Memorial Fellowship**, $3500 per year *2022-2023*

Teaching

- **LING 227/627 Language and Computation I**, Instructor: Kenneth Hanson *Fall 2025, Yale University*
- **LING 384/784 Computational Psycholinguistics**, Instructor: R. Thomas McCoy *Spring 2025, Yale University*
- **LING 380/780 Neural Network Models of Linguistic Structure**, Instructor: Robert Frank *Fall 2024, Yale University*
- **CSE 417 Introduction to Machine Learning**, Instructor: Chien-Ju Ho *Spring 2022, Washington University in St. Louis*
- **CSE 247 Data Structure and Algorithms**, Instructor: Bill Siever *Spring 2021, Washington University in St. Louis*

Reviews

- **CogInterp: First Workshop on Interpreting Cognition in Deep Learning Models** *2025*
- **Association for Computational Linguistics** *2025*

Services

- **Yale Linguistics Department Colloquium Series**, co-organizer *2023-2024*
- **Annual Meeting of the North East Linguistic Society (NELS 55)**, organizing committee member *2024*

Skills

- **Natural Languages:** Mandarin Chinese (native), English (proficient), Japanese (beginner to intermediate)
- **Programming Languages:** Python (proficient), R (proficient), Java (familiar), C++ (familiar)
- **Frameworks:** PyTorch, nnsight, transformerLens, pyvene, NeuroSurgeon