

Xufeng Duan

Email Google Scholar Personal Website

PhD Candidate, [Language Processing Lab](#), The Chinese University of Hong Kong

Research interests: LLM, Mechanistic Interpretability, Neuro-cognitive Mechanism of Handwriting

Education

The Chinese University of Hong Kong Language Processing Lab	2020–2025 (expected) <i>Ph.D.</i> in Linguistics
--	---

The Chinese University of Hong Kong Language Processing Lab	2019–2020 <i>M.A.</i> in Linguistics
--	---

Sun Yat-sen University Neurolinguistics Laboratory	2015–2019 <i>B.A.</i> in Chinese
---	-------------------------------------

Skill

- **LLMs:** fine-tuning, SAE, ablation, benchmarking, etc.
- **Experimentation:** fMRI / TMS / EEG / fNIRS / Eye-tracker using PsychoPy & Experiment Builder & E-prime & Gorilla / Qualtrics.
- **Coding:** Python, R and Matlab.

Publication

Xiao, B., **Duan, X.**, Haslett, D. A., & Cai, Z. G. (in press). Human-likeness of LLMs in the Mental Lexicon. *Conference on Computational Natural Language Learning*.

Qiu, Z., **Duan, X.**, & Cai, Z. G. (2025). Grammaticality representation in ChatGPT as compared to linguists and laypeople. *Humanities and Social Sciences Communications*, 12(1), 1–15. [doi:10.1057/s41599-025-04907-8](https://doi.org/10.1057/s41599-025-04907-8)

Wang, S., **Duan, X.**, & Cai, Z. (2025). What to predict? Exploring how sentence structure influences contrast predictions in humans and large language models. In T. Kuribayashi, G. Rambelli, E. Takmaz, P. Wicke, J. Li, & B.-D. Oh (Eds.), *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 244–252). NAACL 2025. [2025.cmcl-1.28](https://arxiv.org/abs/2025.cmcl-1.28)

Wu, H., **Duan, X.**, & Cai, Z. (2025). Distinct social-linguistic processing between humans and large audio-language models: Evidence from model-brain alignment. In T. Kuribayashi, G. Rambelli, E. Takmaz, P. Wicke, J. Li, & B.-D. Oh (Eds.), *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 135–143). NAACL 2025. [2025.cmcl-1.18](https://arxiv.org/abs/2025.cmcl-1.18)

Duan, X., Li, S., & Cai, Z. G. (2025). MacBehaviour: An R package for behavioural experimentation on large language models. *Behavior Research Methods*, 57(1), 19. [doi:10.3758/s13428-024-02524-y](https://doi.org/10.3758/s13428-024-02524-y)

Duan, X., Zhou, X., Xiao, B., & Cai, Z. (2025). Unveiling language competence neurons: A psycholinguistic approach to model interpretability. In O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, & S. Schockaert (Eds.), *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 10148–10157). Association for Computational Linguistics. <https://aclanthology.org/2025.coling-main.677/>

- Zhou, X., Chen, D., Cahyawijaya, S., **Duan, X.**, & Cai, Z. G. (2025). Linguistic Minimal Pairs Elicit Linguistic Similarity in Large Language Models. *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 6866–6888). Association for Computational Linguistics. <https://aclanthology.org/2025.coling-main.459/>
- Tang, X., **Duan, X.**, & Cai, Z. G. (2024). Are LLMs Good Literature Review Writers? Evaluating the Literature Review Writing Ability of Large Language Models. *arXiv preprint arXiv:2412.13612*. [doi:10.48550/arXiv.2412.13612](https://doi.org/10.48550/arXiv.2412.13612)
- *Wang, S., ***Duan, X.**, & Cai, Z. (2024). A Multimodal Large Language Model “Foresees” Objects Based on Verb Information but Not Gender. *Proceedings of the 28th Conference on Computational Natural Language Learning* (pp. 435–441). [doi:10.18653/v1/2024.conll-1.32](https://doi.org/10.18653/v1/2024.conll-1.32)
- Qiu, Z., **Duan, X.**, & Cai, Z. (2024). Evaluating Grammatical Well-Formedness in Large Language Models: A Comparative Study with Human Judgments. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 189–198). [doi:10.18653/v1/2024.cmcl-1.16](https://doi.org/10.18653/v1/2024.cmcl-1.16)
- Duan, X.**, Xiao, B., Tang, X., & Cai, Z. G. (2024). HLB: Benchmarking LLMs’ Humanlikeness in Language Use. *arXiv:2409.15890 [cs.CL]*. [doi:10.48550/arXiv.2409.15890](https://doi.org/10.48550/arXiv.2409.15890)
- Duan, X.**, & Cai, Z. G. (2024). Chinese Character Processing. *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. [doi:9780199384655.013.1012](https://doi.org/10.1093/acrefore/9780199384655.013.1012)
- Wu, H., **Duan, X.**, & Cai, Z. G. (2024). Speaker Demographics Modulate Listeners’ Neural Correlates of Spoken Word Processing. *Journal of Cognitive Neuroscience*, 36(10), 2208–2226. [doi:10.1162/jocn_a_02225](https://doi.org/10.1162/jocn_a_02225)
- Cai, Z., **Duan, X.**, Haslett, D., Wang, S., & Pickering, M. (2024). Do large language models resemble humans in language use? *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 37–56). [doi:10.18653/v1/2024.cmcl-1.4](https://doi.org/10.18653/v1/2024.cmcl-1.4)
- Qiu, Z., **Duan, X.**, & Cai, Z. (2023). Does ChatGPT Resemble Humans in Processing Implicatures? *Proceedings of the 4th Natural Logic Meets Machine Learning Workshop* (pp. 25–34). <https://aclanthology.org/2023.naloma-1.3>
- Lu, S., **Duan, X.**, & Li, X. (2021). Prominence and Locality Principle in Chinese Possessive Topicalization Sentences and its Psychological Basis—Evidence from eye-movements. *Contemporary Linguistics*, 23(1), 56–73. [\[pdf\]](#)

Conference Talk

- Xufeng Duan** (2024). MacBehaviour: An R package for behavioural experimentation on large language models. *Architectures and Mechanisms for Language Processing Asia (AMLaP Asia)*, Singapore.
- Xufeng Duan** (2024). The Neural Basis of Decline in Written Production: Evidence from Chinese Handwriting. *The annual meeting of Society for the Neurobiology of Language (SNL)*, Brisbane, Australia.
- Xufeng Duan** (2024). Humanlike language use in LLMs. *CogSci Hong Kong Meetup*, Hong Kong, China.
- Xufeng Duan** (2023). Chinese translations of English words: A comparison between human and ChatGPT translations. *Architectures and Mechanisms for Language Processing Asia (AMLaP Asia)*, Hong Kong, China.

- Zhenguang Cai, **Xufeng Duan**, Lu Sun, & Martin Pickering (2023). Structural priming in ChatGPT. *Architectures and Mechanisms for Language Processing Asia (AMLaP Asia)*, Hong Kong, China.
- Zhenguang Cai, **Xufeng Duan**, David Haslett, Shuqi Wang, & Martin Pickering (2023). Do language models resemble humans in language use? *Architectures and Mechanisms for Language Processing (AMLaP)*, Donostia-San Sebastián, Spain.
- Zhuang Qiu, **Xufeng Duan**, & Zhenguang Cai (2023). Does ChatGPT Resemble Humans in Processing Implicatures? *Proceedings of the 4th Natural Logic Meets Machine Learning Workshop*, Nancy, France.
- Xufeng Duan** (2018). Language rehabilitation training therapies for Chinese aphasia — Using Chinese characters as clues. *Beyond A Limited Scope: Chinese for Specific, Academic, & Professional Purposes*, Hawaii, US.
- Xufeng Duan** (2018). The Dependence on different symbolic systems during word recognition-pronunciation tests by Chinese native speakers with aphasia — based on eye-movement methodology. *Annual Conference of the Chinese Society of Neurological Linguistics*, Xuzhou, China.
- Xufeng Duan** (2018). The Rules of Prominency and Locality Psychological Reality Basis. *Annual Academic Conference of the Chinese Linguistic Society*, Guangzhou, China.

Poster Presentation

- Zhang, Y., **Duan, X.**, & Cai, Z. (2024). The causal role of left inferior frontal gyri in Chinese character handwriting: A transcranial magnetic stimulation study. *The annual meeting of Society for the Neurobiology of Language (SNL)*, Brisbane, Australia.
- Qiu, Z., **Duan, X.**, & Cai, Z. (2024). Do Large Language Models Resemble Humans in Grammaticality Judgement? *Human sentence processing 2024*, University of Michigan, USA.
- Duan, X.**, Cai, Z., & Yao, B. (2023). Electrophysiological responses associated with character amnesia in Chinese handwriting. *Architectures and Mechanisms for Language Processing (AMLaP)*, Donostia-San Sebastián, Spain.
- Wu, H., **Duan, X.**, & Cai, Z. (2023). Belief of Speakers' Linguistic Competence Modulates the N400 Effect Elicited by Inconsistent Lexical Use. *Architectures and Mechanisms for Language Processing (AMLaP)*, Donostia-San Sebastián, Spain.
- Qiu, Z., **Duan, X.**, Cai, Z., & Zhao, N. (2023). Pragmatic Implicature Processing in ChatGPT. *Architectures and Mechanisms for Language Processing (AMLaP)*, Donostia-San Sebastián, Spain.

Scholarship and Research Funding

- The Chinese University of Hong Kong Ph.D. Studentship **880,000 HKD, 2020-2024**
- Key Laboratory Open Fund for student **30,000 RMB, 2019**
Multimodal (EEG/fNIRS/Eyetracking) Corpus for Special Population, Sun Yat-sen University.

Training and Internship

-
- **TMS Training** May 2023
APAC TMS clinic Workshop, University of Numnich & MagMore Company.
 - **Language Therapist** Jul. 2017 – Jul. 2019
The Third Affiliated Hospital of Sun Yat-sen University.

Last updated: 24/05/2025