YANG XU

Curriculum Vitae

July 2023

Email: sustech@sustech.edu.cn;
Website: `clcs.sdsu.edu` (old site)
Lab website: under construction.

Department of Computer Science and Engineering
College of Engineering
Southern University of Science and Technology

RESEARCH INTERESTS

Dr. Yang Xu's research interest lies in the intersection of computational linguistics (natural language processing), psycholinguistics, and cognitive sciences. He develops theoretical paths and computational methods for better understand human language, cognition, and the mysterious consciousness of mind. His research draw inspirations from Information-Theory, theories of mind (in psychology), and dynamic complex systems. He is enthusiastic in and has been actively working on the following research questions:

- How natural languages differ from artificial ones (model-generated, e.g., GPTs)?
- How the information in different communication modalities (verbal and non-verbal) distribute and evolve in monologue/dialogue?
- How the dyad of speakers adjust their language to reach better coordination in dialogue and to maintain efficient communication as a dynamical system?
- Connection between language production and the cognitive/physiological constraints.

He is also interested in other broader questions:

- How distributed lexical semantic models reveal historical changes in semantic composition?
- What are the spatial information theoretic patterns of written Chinese language?

PROFESSIONAL EXPERIENCE

Associate Professor   June 2023 – Present
Department of Computer Science and Engineering
Southern University of Science and Technology (SUSTech), Shenzhen, China
Assistant Professor   Aug 2018 – May 2023
Department of Computer Science, San Diego State University (SDSU), CA, US

EDUCATION

Ph.D in Information Sciences and Technology   Aug 2013 – May 2018
The Pennsylvania State University, USA

Master of Art   Sep 2010 – July 2013
Department of Psychology, Tsinghua University, China

Bachelor of Engineering   Sep 2006 – July 2010
Department of Electronic Engineering, Tsinghua University, China

RESEARCH GRANTS

PI, NSF-CRII-HCC (**awarded**), $166K, 2021 – 2023
Modeling Computer-Mediated Task-Oriented Dialogues with Multi-Modality Information Theoretic Approaches
Award ID: 2105192

Co-PI, NIH (subaward to SDSU HealthLink) (**awarded**), $50K, 2021 – 2022
An NLP Approach to Analyze the Past and Present, and Inform the Future of Health Disparities Research

PI, Division of Research and Innovation, SDSU(**awarded**), $100K, 2021 – 2022
Supporting Deep Learning-based Research Activities with GPU-equipped High Performance Server Computers

Co-PI, NSF-SCH, $1.4M (pending), 2021
Examine Your Eye: Smart Translation of Erythrocyte moving pattern for iLLness prediction (EYES TELL)

PI, NSF-RETTL, $800K (unfunded), 2021
Enabling Interactive Online Education based on Multimodal Analysis and Multi-Modality Machine Learning Techniques

PI, NSF-CRII, $200K (unfunded), 2019
Towards Explainable Word Embeddings: A Historical Language Change Perspective

PI, NSF-IIS, $500K (unfunded), 2018
Computational Studies of Long-term Language Change by Fine-tuning Word Embedding Models

PUBLICATIONS

Published

1. **Yang Xu** and Yang Cheng. Spontaneous gestures encoded by hand positions improve language models: An information-theoretic motivated study. 2023. (To appear in ACL 2023; CCF-A)

2. **Yang Xu**, Yang Cheng, and Riya Bhatia. Gestures are used rationally: Information theoretic evidence from neural sequential models. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 134–140, Gyeongju, Republic of Korea, 2022. International Committee on Computational Linguistics. (CCF-B)

3. Bruce I. Gaynes, Mark B. Shapiro, Jie Xu, Abel Saju Augustine, **Yang Xu**, and et.al. Hierarchical data visualization of experimental erythrocyte aggregation employing cross correlation and optical flow applications. *Microvascular Research*, 143:104386, 2022

4. **Yang Xu**. Global divergence and local convergence of utterance semantic representations in dialogue. In *Proceedings of the Society for Computation in Linguistics*, volume 4, pages 116–124. 2021

5. **Yang Xu** and Zheng-sheng Zhang. Historical changes in semantic weights of sub-word units. In Adam Jatowt, Nina Tahmasebi, Yang Xu, and Lars Borin, editors, *Computational Approaches to Semantic Change*, pages 169–187. Language Science Press, 2021

6. M Alex Kelly, **Yang Xu**, Jesús Calvillo, and David Reitter. Which sentence embeddings and which layers encode syntactic structure? In *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*, pages 2375–2381, 2020. (acceptance rate: 18.1%; CCF-B)

7. **Yang Xu**, Jiasheng Zhang, and David Reitter. Treat the word as a whole or look inside? subword embeddings model language change and typology. In *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, pages 136–145, Florence, Italy, 2019. Association for Computational Linguistics

8. **Yang Xu**, Jeremy Cole, and David Reitter. Linguistic alignment is affected more by lexical surprisal rather than social power. *Proceedings of the Society for Computation in Linguistics*, 2(1):349–352, 2019

9. **Yang Xu**, Jeremy Cole, and David Reitter. Not that much power: Linguistic alignment is influenced more by low-level linguistic features rather than social power. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 601–610, 2018. (acceptance rate: 24.5%; CCF-A)

10. **Yang Xu** and David Reitter. Information density converges in dialogue: Towards an information-theoretic model. *Cognition*, 170:147–163, 2018. (Impact factor: 3.65; CCF-B)

11. **Yang Xu** and David Reitter. Spectral analysis of information density in dialogue predicts collaborative task performance. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 623–633, Vancouver, Canada, 2017. (acceptance rate: 25%; CCF-A)

12. **Yang Xu** and David Reitter. Entropy converges between dialogue participants: Explanations from an information-theoretic perspective. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 537–546, Berlin, Germany, August 2016. Association for Computational Linguistics. (acceptance rate: 28.0%; CCF-A)

13. **Yang Xu** and David Reitter. Convergence of syntactic complexity in conversation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 443–448, Berlin, Germany, August 2016. Association for Computational Linguistics. (acceptance rate: 28.0%; CCF-A)

14. **Yang Xu** and David Reitter. An evaluation and comparison of linguistic alignment measures. In *Proceedings of Cognitive Modeling and Computational Linguistics (CMCL)*, pages 58–67, Denver, CO, 2015. Association for Computational Linguistics

15. Alexander G. Ororbia II, **Yang Xu**, Vito D'Orazio, and David Reitter. Error-correction and aggregation in crowd-sourcing of geopolitical incident information. In N. Agarwal et al., editor, *Social Computing, Behavioral Modeling and Prediction*, volume 9021 of *Lecture Notes in Computer Science*, pages 381–387. Springer, 2015

16. David Reitter, **Yang Xu**, Patrick Craven, Anik Sndor, R. Chris Garrett, E. Vince Cross, and Jerry L. Franke. Cognitive models predicting surprise in robot operators. In *Proc. International Conference on Cognitive Modeling*, pages 190–191, Groningen, Netherlands, 2015

17. **Yang Xu** and Hong Li. Influence of visibility range and degree of urgency on efficiency of evacuation: Mediating effect of herding behavior. *Studies of Psychology and Behavior*, 13(3):311–319, 2015. (Published in Chinese)

18. Hong Li, **Yang Xu**, Shi Chen, and Anqi Gao. Effects of intuition and analysis on the adoption of sources of information in escape decision-making. *Studies of Psychology and Behavior*, 10(6):452–458, 2012. (Published in Chinese)

CONFERENCE PRESENTATIONS (SELECTED)

1. Global Divergence and Local Convergence of Dialogue, virtual, *SCiL*, 2021

2. Subword Embeddings Reveal Language Change, poster, *CogSci*, Montreal, Canada, 2019

3. Treat the Word As a Whole or Look Inside, poster, *ACL*, Florence, Italy, 2019

INVITED TALKS

University of Waterloo, Canada   Aug 2022
Host: Dr. Moojan Ghafurian
Title: Speaking Rationally by Gestures: Evidence from Multi-Modal Language Models

Beijing Language and Culture University   Aug 2021
Host: Dr. Qun Guan
Title: The Information-Theoretic Insights of Multimodal Language Models

Beijing Language and Culture University   Aug 2020
Host: Dr. Qun Guan
Title: Entropy Constancy: Insights about Natural Language from the Perspective of Information

Westlake University, China   Nov 2019
Host: NLP Group of Dr. Yue Zhang   *Title*: Dialogue from an Information-Theoretic Perspective

Shanghai Jiao-Tong University, China   Nov 2019
Host: College of Engineering   *Title*: Efficient Communication in Dialogue

Tsinghua University, China   Nov 2019
Host: Dept. of Psychology   *Title*: An Information Theoretic Psycholinguistic Model of Dialogue

San Diego State University   Oct 2019
Host: Dept. of Math and Stats   *Title*: Statistical Methods Used in Natural Language Processing

San Diego State University   Oct 2019
Host: SDSU A.I. Seminar   *Title*: Recent Advances in Natural Language Processing

San Diego State University   Feb 2018
Host: Dept. of CS   *Title*: Computational Understanding of Dialogue and Language

HONOR AND AWARDS

Award for Research Excellence	Penn State, College of Information Sciences and Technology, 2018
Outstanding Graduates Award	Tsinghua University, 2013
Distinguished Master Thesis Award	Tsinghua University, 2013
Fellowship of Excellence in College Entrance Exam	Changxing, Zhejiang, 2006

PROFESSIONAL SERVICES

Reviewer for ACL 2019, 2020, 2021, 2023
Reviewer for CONLL 2019, 2020, 2021, 2022, 2023
Reviewer for CogSci 2018, 2019, 2020

Reviewer for *Lingua*	2022
Reviewer for *IEEE Transactions on Learning Technologies*	2022
Reviewer for *Dialogue and Discourse*	2022
Reviewer for *Journal of Psycholinguistic Research*	2021
Reviewer for *Technology, Mind, and Behavior*	2020

COURSES TAUGHT

CS460 Algorithms	2018 − 2022
CS549 Machine Learning	2018 − 2022