

---

## Professional Positions

- 08/2023- **Assistant Professor**, School of Languages and Cultures, Purdue University
- 08/2023- **Affiliate**, School of Interdisciplinary Studies, Purdue University
- 01/2024- **Faculty Associate**, Center on Aging and the Life Course, Purdue University
- 2021-2023 **Postdoctoral Fellow**, Feinstein Institutes, Northwell Health
- 2015-2016 **Project Assistant**, The Hong Kong Polytechnic University

---

## Education

- 07/2021 **Ph.D. in Linguistics**, Michigan State University
- 02/2015 **M.A. in Language Studies**, The Hong Kong Baptist University

---

## Peer-Reviewed Journal Articles

- 14. Cho, Sunghye; **Cong, Yan**; Mehta, Aarush; Nikzad, Amir H.; Berretta, Sarah A.; Behbehani, Leily M.; Liberman, Mark; Tang, Sunny X.. (2026). Unique signatures in verbal fluency task performance in schizophrenia and depression. *Schizophrenia Research: Cognition*, 43, 100407.
- 13. **Cong, Yan** and Lee, Jiyeon. (2025). Tracking priming-induced language recovery in aphasia with pre-trained language models. *Frontiers in Artificial Intelligence*, 8, 1668399.
- 12. Mehta, Aarush; Nikzad, Amir; **Cong, Yan**; Cho, Sunghye; Pradhan, Sameer; Tang, Sunny. X. (2025). Sentiment in speech is associated with symptom severity in psychosis. *Cognitive Neuropsychiatry*. 1–12.
- 11. **Cong, Yan**. (2025). Second language learning of degree expressions: A computational approach. *Cambridge Natural Language Processing*. 31(5), 1187–1209.
- 10. **Cong, Yan**. (2024). Manner implicatures in large language models. *Nature Scientific Reports*. 14, 29113.
- 9. Britton, James; **Cong, Yan**; Hsu, Yu-Yin; Chersoni, Emmanuele; and Blache, Philippe. (2024). On the Influence of Discourse Connectives on the Predictions of Humans and Language Models. *Frontiers in Human Neuroscience*. 18, 1363120.
- 8. **Cong, Yan**. (2024). Demystifying large language models in second language development research. *Computer Speech & Language*. 89, 101700.
- 7. **Cong, Yan**. (2024). AI Language Models: An Opportunity to Enhance Language Learning. *Informatics*. 11(3), 49.

6. **Cong, Yan**; LaCroix, Arianna N.; Lee, Jiyeon. (2024). Clinical efficacy of pre-trained large language models through the lens of aphasia. *Nature Scientific Reports*. 14, 15573.
5. Tang, Sunny X.; **Cong, Yan**; Mercep, Gwentyth; Bhatti, Mutahira; Serpe, Grace; Gromova, Valeria; Berretta, Sarah; John, Majnu; Liberman, Mark; Sinvani, Liron. (2023). Characterizing and Detecting Delirium with Clinical and Computational measures of Speech and Language Disturbance. *Journal of Psychiatry and Neuroscience*. 48(4), E255-E264.
4. Tang, Sunny X.; **Cong, Yan**; Nikzad, Amir; Mehta, Aarush; Cho, Sunghye; Hänsel, Katrin; Beretta, Sarah; Dhar, Aamina; Kane, John; Malhotra, Anil. (2022). Clinical and Computational Speech Measures are Associated with Social Cognition in Schizophrenia Spectrum Disorders. *Schizophrenia research*. 259, 28-37.
3. Nikzad, Amir; **Cong, Yan**; Beretta, Sarah; Hänsel, Katrin; Cho, Sunghye; Pradhan, Sameer; Behbehani, Leily; DeSouza, Danielle; Liberman, Mark; Tang, Sunny X. (2022). Who Does What to Whom? Graph Representations of Action-Predication in Speech Relate to Psychopathological Dimensions of Psychosis. *Nature Schizophrenia*. 8, 58.
2. Tang, Sunny X.; Hänsel, Katrin; **Cong, Yan**; Nikzad, Amir; Cho, Sunghye; Berretta, Sarah; Behbehani, Leily. (2022). Latent Factors of Speech and Language Disturbance and Relationships to Acoustic and Lexical Computational Features. *Schizophrenia Bulletin*. 49(Supplement\_2), S93-S103.
1. **Cong, Yan**. (2022). Judge-dependence in Quality Nouns: A Semantic Analysis of the Mandarin Chinese “you NP” Structure. *Asian Language and Linguistics*. 2(2), 167-216.

---

## Peer-Reviewed Conference Proceedings Articles

13. **Cong, Yan**; Rayz, Julia. (2025). Language models demonstrate the good-enough processing seen in humans. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. 47(0):4642-4649.
12. Li, Yue; **Cong, Yan**; Francis, Elaine J. (2025). Beyond Binary Animacy: A Multi-Method Investigation of LMs’ Sensitivity in English Object Relative Clauses. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*. 184–196. Association for Computational Linguistics (ACL).
11. Hu, Jingying; **Cong, Yan**. (2025). Modeling Chinese L2 Writing Development: The LLM-Surprisal Perspective. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*. 172–183. Association for Computational Linguistics (ACL).
10. **Cong, Yan**; Lee, Jiyeon; LaCroix, Arianna N. (2024). Leveraging pre-trained large language models for aphasia detection in English and Chinese speakers. In *Proceedings of the 6th Clinical Natural Language Processing Workshop (Clinical NLP)*. 238-245. Association for Computational Linguistics (ACL).

9. A, Pranav; **Cong, Yan**; Chersoni, Emmanuele; Hsu, Yu-Yin.; Lenci, Alessandro. (2024). Comparing Static and Contextual Distributional Semantic Models on Intrinsic Tasks: An Evaluation on Mandarin Chinese Datasets. In *Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)*. 3610-3627. Association for Computational Linguistics (ACL).
8. **Cong, Yan**; Chersoni, Emmanuele; Hsu, Yu-Yin; Blache, Philippe. (2023). Investigating the Effect of Discourse Connectives on Transformer Surprisal: Language Models Understand Connectives, Even So They Are Surprised. In *Proceedings of the 6th Conference on Blackbox NLP*. 222-232. Association for Computational Linguistics (ACL).
7. **Cong, Yan**; Chersoni, Emmanuele; Hsu, Yu-Yin; Lenci, Alessandro. (2023). Are Language Models Sensitive to Semantic Attraction? A Study on Surprisal. In *Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (\*SEM)*. 141-148. Association for Computational Linguistics (ACL).
6. **Cong, Yan**. (2022). Psycholinguistic Diagnosis of Language Models' Commonsense Reasoning. In *Proceedings of the First Workshop on Commonsense Representation and Reasoning (CSRR)*. 17-22. Association for Computational Linguistics (ACL).
5. **Cong, Yan**. (2022). Pre-trained Language Models' Interpretation of Evaluativity Implicature: Evidence from Gradable Adjectives Usage in Context. In *Proceedings of the Second Workshop on Understanding Implicit and Underspecified Language*. 1-7. Association for Computational Linguistics (ACL).
4. **Cong, Yan**; Ngonyani, Deo. (2022). A syntactic analysis of the co-occurrence of stative and passive in Kiswahili. In Sibanda, Galen; Ngonyani, Deo; Choti, Jonathan; Biersteker, Ann (eds.). *Descriptive and theoretical approaches to African linguistics*. 165-181.
3. Pandia, Lalchand; **Cong, Yan**; Ettinger, Allyson. (2021). Pragmatic competence of pre-trained language models through the lens of discourse connectives. In *Proceedings of the 25th Conference on Computational Natural Language Learning (CoNLL)*. 367-379. Association for Computational Linguistics (ACL).
2. **Cong, Yan**. (2019). The Non-canonical Argument Realization of Chinese Alternations. In *Proceedings of the 30th North American Conference on Chinese Linguistics (NACCL)*. 1, 82-99.
1. **Cong, Yan**; Wee, Lian-Hee; Huang, Chu-Ren. (2015). The Invertible Construction in Chinese. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation (PACLIC)*. 206-211. Association for Computational Linguistics (ACL).

---

## Peer-Reviewed Conference Abstracts

27. **Cong, Yan** and Lee, Jiyeon. (2025). *Language models derived surprisal and structural priming in aphasia: Implications for language recovery*. In Academy of Aphasia Annual Meeting 2025. October 26 - 28, 2025; San Diego, CA.

26. Fang, Shaohua; Li, Yue; **Cong, Yan**. (2025). *Understanding Quantifier Scope with Large Language Models: How Many Children Climbed Trees?*. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 47). July 29 - August 1, 2025; San Francisco, CA.
25. **Cong, Yan**; LaCroix, Arianna N.; Jiyeon Lee. (2025). *Voice AI Symposium, Bridge2AI-Voice Consortium*. Enhancing Stroke-Induced Aphasia Assessment Using Interpretable linguistic biomarkers and Large Language Models. April 22-24, 2025; University of South Florida, CA.
24. Nikzad , Amir H.; **Cong, Yan**; Cho, Sunghye; Partlan, Ryan; Li, Jiefei; and Tang, Sunny X. (2025). *The 2025 Congress of the Schizophrenia International Research Society*. Evaluating Language Models and Preprocessing Strategies for Detecting Psychosis-Related Language Disturbances. March 29 - April 2, 2025; Chicago, IL.
23. Hu, Jingying and **Cong, Yan**. (2025). *The 8th International Forum on Linguistics and Chinese Education (IFOLCE) & The 32nd International Conference on Chinese Language Instruction*. Leveraging AI language models for Chinese L2 Writing Assessment. April 5-6, 2025; Princeton University, NJ.
22. Issa, Elsayed and **Cong, Yan**. (2025). *CALICO (Computer Assisted Language Instruction Consortium) Annual Conference*. Elicited Imitation Revisited: an open-sourced package for assessing L2 speech. May 27-31, 2025; San Diego State University, CA.
21. Fang, Shaohua; Li, Yue; **Cong, Yan**. (2025). *CLTA (The Chinese Language Teachers Association, USA, Inc.) Annual Conference*. Understanding Quantifier Scope with Large Language Models: How Many Children Climbed Trees? May 9-10, 2025; New York, NY.
20. Gautam, Sukrati; Yu, David J; **Cong, Yan**. (2024). *American Geophysical Union (AGU) Annual Meeting*. Enhancing Resilience in Reservoir Management: Addressing Logical Interdependencies through Framework Using Large Language Models and Natural Language Processing. December 9-13, 2024; Washington, D.C.
19. Issa, Elsayed; **Cong, Yan**; Scala, Carmela; Merschel, Lisa; Frattarola, Angela. Organizers: William, Jennifer; Fellin, Luciana. (2024). *Modern Language Association (MLA) Convention*. AI, NLP, and Computational Linguistics: Implications for Language Programs. January 4-7, 2024; Philadelphia, PA.
18. Nikzad, Amir H.; **Cong, Yan**; Berretta, Sarah; Behbehani, Leily; Pradhan, Sameer; Tang, Sunny X. (2023). *The 62nd Annual Meeting of the American College of Neuropsychopharmacology*. Disorganization Disrupts the Core: Localizing Clinical Dimensions of Speech and Language Disturbance Within Clause Structure Using Multi-Layered Semantic Graph Analysis. December 3-6, 2023; Tampa, FL.

17. Tang, Sunny X.; Nikzad, Amir H.; Spilka, MJ; **Cong, Yan**; Behbehani, Leily; Berretta, Sarah; Kane, JM; Malhotra, AK. (2023). *The 62nd Annual Meeting of the American College of Neuropsychopharmacology*. Linguistic risk score for schizophrenia: applying a polygenic risk approach to synthesize multivariate features from computational speech and language analysis. December 3-6, 2023; Tampa, FL.
16. **Cong, Yan**; Wolff, Phillip. (2022). *Annual Meeting of The Linguistic Society of America (LSA)*. Inferring markedness from semantic weight: An approach using the T5 language model. Washington, D.C.
15. Nikzad, Amir; **Cong, Yan**; Hänsel, Katrin; Cho, Sunghye; Berretta, Sarah; Behbehani, Leily; Tang, Sunny X. (2022). *Annual Meeting of The Society of Biological Psychiatry (SOBP)*. Modeling cross-diagnostic speech disturbance with semantic and structural speech graphs. Virtual presentation.
14. Hänsel, Katrin; Nikzad, Amir; **Cong, Yan**; Mehta, Aarush; Cho, Sunghye; Berretta, Sarah; Behbehani, Leily; Tang, Sunny X. (2022). *Annual Meeting of The Society of Biological Psychiatry (SOBP)*. Predicting general psychopathology with multi-modal speech and language features. Virtual presentation.
13. **Cong, Yan** and Tang, Sunny X. (2022). *The 6th Annual Technology in Psychiatry Summit*. Leveraging state-of-the-art transformer language models to measure speech disturbance in schizophrenia. Harvard Medical School, Mclean Hospital.
12. **Cong, Yan**; Buccola, Brian. (2021). *The 45th annual Penn Linguistics Conference (PLC)*. Alternatives and symmetry in non-scalar implicature: The case of Mandarin adjective constructions. Philadelphia, Pennsylvania.
11. Hänsel, Katrin; Cho, Sunghye; Berretta, Sarah; Mehta, Aarush; Nikzad, Amir; Guo, Jessica; **Cong, Yan**; Pradhan, Sameer; Fiumara, James; Liberman, Mark; Tang, Sunny X. (2021). *The 60th Annual meeting of American College of Neuropsychopharmacology (ACNP)*. Associative dysregulation and impaired expressivity in schizophrenia: Latent structures of clinical and automated speech analysis features. Virtual presentation.
10. Smith, Kaylin; Ma, Ye; **Cong, Yan**; Beretta, Alan. (2019). *The Psycholinguistics in Iceland—Parsing and Prediction (PiPP)*. Interactions between syntax and semantics: The semantic P600, University of Iceland. Reykjavík, Iceland.
9. Smith, Kaylin; Ma, Ye; **Cong, Yan**; Beretta, Alan. (2019). *The 32nd CUNY Conference on Human Sentence Processing*. Semantic attraction in sentence processing. University of Colorado Boulder. Boulder, Colorado.
8. **Cong, Yan**. (2019). *The Linguistic Society of America Linguistics Institute (LSA)*. Information structure and the depth of language processing: a pilot study. University of California at Davis. California.
7. **Cong, Yan**. (2019). *International Workshop on Degrees and Grammar: An East Asian Perspective*. A subjective analysis of the Mandarin ‘you’ structure. Nanjing University, China.

6. **Cong, Yan.** (2018). *The Annual Conference of the Chinese Language Teachers Association (CLTA)*. The non-canonical argument realization of the Chinese alternations. Washington, D.C.
5. **Cong, Yan.** (2018). *The 12th Arizona Linguistics Circle (ALC)*. Rethinking the implicit argument in Kiswahili. University of Arizona. Tucson, Arizona.
4. **Cong, Yan.** (2018). *The 2nd Crete Linguistics Summer School (CreteLing)*. A preliminary semantic analysis of predicates of taste in Chinese. University of Crete. Greece.
3. Parrish, Alicia; Kelley, Patrick; Smith, Kaylin; **Cong, Yan**; Beretta, Alan. (2017). *Annual Meeting of The Society for the Neurobiology of Language (SNL)*. Prosodic lengthening and boundary prediction in nominal compounds: An ERP study. Baltimore, Maryland.
2. **Cong, Yan.** (2017). *The Linguistic Society of America Linguistics Institute (LSA)*. T1 and T0 Sandhi in Tongliao Mandarin: an OT approach. University of Kentucky. Lexington, Kentucky.
1. **Cong, Yan.** (2016). *The 28th North American Conference on Chinese Linguistics (NACCL)*. Markedness theory and the acquisition of the Mandarin potential complement construction. Brigham Young University. Salt Lake City, Utah.

---

## Invited Talks

- 2024 **Cong, Yan.** *Natural language processing and its implications in medicine and education*. Purdue System Thinkers (PurSysT) Research Colloquiums. Purdue University.
- 2024 **Cong, Yan.** *Language Understanding in Humans and AI*. Symposium: Current Ideas in Artificial Intelligence. The 50th Annual Meeting of the Society for Philosophy & Psychology. Purdue University.
- 2024 **Cong, Yan.** *Computational approach to linguistic biomarkers in aging*. Colloquium series. Center on Aging and the Life Course (CALC), Purdue University.
- 2023 **Cong, Yan.** *Large language models and their implications in text analysis*. Round table on *STEM in Asia*. Department of History, Purdue University.

---

## Other Talks

- 2025 William, Jen; Issa, Elsayed; **Cong, Yan.** *Human side of AI*. BASF. Purdue University.
- 2025 **Cong, Yan.** *Humans in Charge, AI in the Loop*. Cornerstone retreat. College of Liberal Arts. Purdue University.
- 2025 **Cong, Yan.** *Reasoning with Enriched Meaning*. SLC colloquium. School of Languages and Cultures. Purdue University.
- 2024 **Cong, Yan**; Issa, Elsayed; Rayz, Julia T. *Leveraging AI in Language Learning*. College of Liberal Arts Research Academy (CLARA) Research Day. Purdue University.

- 2023 **Cong, Yan.** *Exploring L2 Writing Proficiency via LLMs Metrics: a Preliminary Study.* LINGTalk. Department of Linguistics, Purdue University.
- 2023 **Cong, Yan.** *Using large language models to measure fluency.* Advanced Methods at Purdue (AMAP) in the behavioral, health, and social sciences. Purdue University.
- 2023 **Cong, Yan;** Nikzad, Amir; Tang, Sunny X. *Multi-modal computational speech and language features for assessing psychiatric conditions.* The University of Maryland.
- 2023 **Cong, Yan;** Hänsel, Katrin; Nikzad, Amir; Tang, Sunny X. *Investigating novel NLP approaches for measuring speech disturbances in psychosis.* The Functional Neuroimaging and Bioinformatics Lab. McLean, Harvard Medical School Affiliate.
- 2021 **Cong, Yan.** *Formal Linguistics and Philosophy (FLaP).* What we don't say matters. The Hong Kong Baptist University, Department of English.

---

## Fellowships, Funds, and Awards

- 2025 **John Martinson Honors College, Research Breakthrough Award,** *Developing self-managed story-based treatment for aphasia and AI-assisted predictors for language recovery,* Purdue University. PI: Jiyeon Lee. Co-PI: Catherine Pham and Yan Cong. \$60,000
- 2024 **Systems Fellows Award,** Purdue Systems Collaboratory, Purdue University.
- 2024 **Internal fund,** Faculty Research Development Award.  
*Leveraging large language models for aphasia detection and prediction in older adults.* Center on Aging and the Life Course (CALC), Purdue University. \$10,000
- 2023 **Educational Advancement Award**  
*Computational approach to speech biomarkers in bilingual immigrants with psychosis.* *Advancing Women in Science and Medicine,* Northwell health. \$5,000
- 2022 **Educational Advancement Award**  
*Computational linguistics can predict psychiatric diagnoses and outcome.* *Advancing Women in Science and Medicine,* Northwell health. \$5,000
- 2020 **Dissertation Completion Fellowship**  
MSU Graduate School \$10,000
- 2018 **Student Fellowship,** *North American Summer School on Logic, Language, and Information (NASSLLI),* at Pittsburgh, Pennsylvania  
Carnegie Mellon University \$500
- 2017-2020 **Summer Support Fellowship**  
MSU Department of Linguistics and Languages \$4,500
- 2015 **Best Paper Award,** *The 29th Pacific Asia Conference on Language, Information and Computation,* The invertible construction in Chinese. Shanghai Jiaotong University, China

---

## Teaching



- Undergraduate
- LC490-LING398: Language Learning and AI, Purdue (2024-present)
  - SCLA102: Cornerstone - Critical thinking and communication, Purdue (2023-present)
  - LIN401: Introduction to Linguistics, MSU (2017-2021)
  - CHS101, CHS102, CHS411: Chinese language, MSU (2016-2017)
- Graduate
- LING598-LC596: Statistical Modeling in Linguistics II, Purdue (2025-present)
  - LING598-LC596: Statistical Modeling in Linguistics, Purdue (2023-present)

---

## Advising

- Ph.D committee member
- SLC / Linguistics: Yue Li; Jingying Hu; Daeun Kang; Karen Fullin.
  - Computer and Information Technology: Yifei Hu (graduated in fall 2024); Damini Zhang.
  - Civil Engineering: Sukrati Gautam (Purdue Systems Collaboratory Fellow advisor)
  - Speech and Language Pathology: Peng Zhang.

### Undergraduate Involvement in research

- Emily A. Tumacder: Presented at the Spring 2025 Undergraduate Research Conference: Engineering, April 08.
- Cameron Pilla: Machine learning, large language models deployment
- Trisha Godara: NLP pipelines development and optimization, vision and language models

---

## Service to the Profession

- 2025-present Guest Editor: Large language models in psychological science. *Nature Humanities and Social Sciences Communications*.
- 2024-present Editorial Board Member: *Humanities and Social Sciences Communications*, *Springer Nature portfolio* (HSS).
- 2024-present Program Committee and Reviewer:
- *The 2026 International Conference on Language Resources and Evaluation* (LREC 2026).
  - *The 39th Pacific Asia Conference on Language, Information and Computation* (PACLIC 2025).
  - *The 31st International Conference on Computational Linguistics* (COLING 2024).
  - *COLM 2025 Workshop Pragmatic Reasoning in Language Models* (PragLM).



- 2022-present Reviewer (Ad hoc):
- *The Association for Computational Linguistics* (ACL Rolling Review)
  - *Language learning*
  - *Research Methods in Applied Linguistics*
  - *Nature Communications*
  - *Nature Portfolio*
  - *Nature Scientific Reports*
  - *Nature Communications Medicine*
  - *Advanced science*
  - *PLOS One*
  - *Association of Contemporary African Linguistics* (ACAL)
  - *American Journal of Speech-Language Pathology* (AJSLP)
  - *npj Mental Health*
  - *Journal of Speech, Language, and Hearing Research* (JSLHR)
  - *Health Informatics Journal*
  - *Journal of Medical Internet Research*

---

## Service to Purdue University

- |           |  |                 |
|-----------|--|-----------------|
| Member    | Graduate admissions subcommittee.                      | since Fall 2024 |
| Member    | CLA Ross-Lynn Postdoctoral Fellow review subcommittee. | for Fall 2024   |
| Organizer | Purdue Computational Linguistics Lab.                  | since Fall 2023 |

---

## Non-Degree Education

- |      |   |
|------|---|
| 2020 | North American Summer School for Logic, Language, and Information (NASSLLI), Brandeis University (Online due to COVID-19) |
| 2019 | Linguistics Society of America Summer Institute (LSA), University of California, Davis                                    |
| 2018 | Crete Summer School in Linguistics, University of Crete   |
| 2018 | NASSLLI, Carnegie Mellon University   |
| 2017 | LSA, University of Kentucky   |

December 1, 2025