Yan Cong

📱 *+1 (517) 940 1452*
✉ *yancong222@gmail.com*
🌐 `https://yancong222.github.io/`

Current Employment

08/2023-
present
School of Languages and Cultures, Purdue University, *Assistant Professor*
Linguistics, NLP, Chinese linguistics

Academic and Professional Qualifications

08/2021-
07/2023
Postdoctoral fellow, *Feinstein Institutes, Northwell Health*
Natural Language Processing, Speech and Language Pathology

07/2021 **Ph.D. in Linguistics, with Cognitive Science Specialization**, *Michigan State University*
Dissertation: Competition in Natural Language Meaning - the Case of Adjectival Constructions

02/2015 **M.A. in Language Studies, emphasis in Second Language Acquisition**, *Hong Kong Baptist University* , Graduated with merit
Dissertation: The Second Language Acquisition of the Mandarin Potential Complement Construction

Journal Articles and Book Chapters

2023 Tang, Sunny; **Cong, Yan**; Mercep, Gwenyth; Bhatti, Mutahira; Serpe, Grace; Gromova, Valeria; Berretta, Sarah; John, Majnu; Liberman, Mark; Sinvani, Liron. *Journal of Psychiatry and Neuroscience*. Characterizing and Detecting Delirium with Clinical and Computational measures of Speech and Language Disturbance. `https://doi.org/10.1503/jpn.230026`

2022 **Cong, Yan**. *Asian Language and Linguistics* 2:2, pp.167-216. Judge-dependence in Quality Nouns: A Semantic Analysis of the Mandarin Chinese "*you* NP" Structure. `https://doi.org/10.1075/alal.21007.con`

2022 Tang, Sunny; **Cong, Yan**; Nikzad, Amir; Mehta, Aarush; Cho, Sunghye; Hänsel, Katrin; Beretta, Sarah; Dhar, Aamina; Kane, John; Malhotra, Anil. *Special Issue of Schizophrenia Research*. Clinical and Computational Speech Measures are Associated with Social Cognition in Schizophrenia Spectrum Disorders. *Schizophrenia research*. `https://doi.org/10.1016/j.schres.2022.06.012`

2022 Nikzad, Amir; **Cong, Yan**; Beretta, Sarah; Hänsel, Katrin; Cho, Sunghye; Pradhan, Sameer; Behbehani, Leily; DeSouza, Danielle; Liberman, Mark; Tang, Sunny. *Nature Schizophrenia*. Who Does What to Whom? Graph Representations of Action-Predication in Speech Relate to Psychopathological Dimensions of Psychosis. *Schizophrenia 8, 58*. `https://doi.org/10.1038/s41537-022-00263-7`

2022 **Cong, Yan**; Ngonyani, Deo. 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. Berlin: Language Science Press. DOI: 10.5281/zenodo.6393748. `https://langsci-press.org/catalog/book/306`

2022   Tang, Sunny; Hänsel, Katrin; **Cong, Yan**; Nikzad, Amir; Cho, Sunghye; Berretta, Sarah; Behbehani, Leily. Latent Factors of Speech and Language Disturbance and Relationships to Acoustic and Lexical Computational Features. _Special Issue of Schizophrenia Bulletin_. `https://doi.org/10.1093/schbul/sbac145`

2021   **Cong, Yan**. _Competition in Natural Language Meaning: The Case of Adjective Constructions in Mandarin Chinese and Beyond_. Doctoral dissertation, Michigan State University. `https://doi.org/doi:10.25335/z1cs-7363`

Peer-reviewed Proceedings Papers

2023   **Cong, Yan**; Chersoni, Emmanuele; Hsu, Yu-Yin; Blache, Philippe. _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. `To-Appear`

2023   **Cong, Yan**; Chersoni, Emmanuele; Hsu, Yu-Yin; Lenci, Alessandro. _Are Language Models Sensitive to Semantic Attraction? A Study on Surprisal_. In Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (*SEM 2023), pages 141-148, Association for Computational Linguistics. `https://aclanthology.org/2023.starsem-1.13.pdf`

2022   **Cong, Yan**. _Psycholinguistic Diagnosis of Language Models' Commonsense Reasoning_. In Proceedings of the First Workshop on Commonsense Representation and Reasoning (CSRR 2022), pages 17–22, Dublin, Ireland. Association for Computational Linguistics. `https://aclanthology.org/2022.csrr-1.3/`

2022   **Cong, Yan**. _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, pages 1–7, Seattle, USA. Association for Computational Linguistics. `https://aclanthology.org/2022.unimplicit-1.1/`

2021   Pandia, Lalchand; **Cong, Yan**; Ettinger, Allyson. _Pragmatic competence of pre-trained language models through the lens of discourse connectives_. In Proceedings of the 25th Conference on Computational Natural Language Learning, pages 367–379, Online. Association for Computational Linguistics. `https://aclanthology.org/2021.conll-1.29/`

2019   **Cong, Yan**. _The Non-canonical Argument Realization of Chinese Alternations_. In Proceedings of the 30th North American Conference on Chinese Linguistics (NACCL-30). Volume 1. Pages 82-99. `https://www.researchgate.net/publication/341598112_The_Non-canonical_Argument_Realization_of_Chinese_Alternations`

2015   **Cong, Yan**; Wee, Lian-Hee; Huang, Chu-Ren. _The Invertible Construction in Chinese_. In Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, pages 206–211, Shanghai, China. `https://aclanthology.org/Y15-1024/`

Invited talks

2023   **Cong, Yan**. _Large language models and their implications in text analysis_. Round table on _STEM in Asia_. 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.

2022 **Cong, Yan** and Tang, Sunny X. *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.

2021 **Cong, Yan**. *Formal Linguistics and Philosophy (FLaP)*. What we don't say matters. The Hong Kong Baptist University, Department of English.

Selected Presentations and Abstracts at Peer-Reviewed Conferences

2022 **Cong, Yan**; Wolff, Phillip. *Annual Meeting of The Linguistic Society of America (LSA)*. Inferring markedness from semantic weight: An approach using the T5 language model. Washington, D.C.

2022 Nikzad, Amir; **Cong, Yan**; Hänsel, Katrin; Cho, Sunghye; Berretta, Sarah; Behbehani, Leily; Tang, Sunny. *Annual Meeting of The Society of Biological Psychiatry (SOBP)*. Modeling cross-diagnostic speech disturbance with semantic and structural speech graphs. Virtual presentation.

2022 Hänsel, Katrin; Nikzad, Amir; **Cong, Yan**; Mehta, Aarush; Cho, Sunghye; Berretta, Sarah; Behbehani, Leily; Tang, Sunny. *Annual Meeting of The Society of Biological Psychiatry (SOBP)*. Predicting general psychopathology with multi-modal speech and language features. Virtual presentation.

2021 **Cong, Yan**; Buccola, Brian. *The 45th annual Penn Linguistics Conference (PLC)*. Alternatives and symmetry in non-scalar implicature: The case of Mandarin adjective constructions. Philadelphia, Pennsylvania.

2021 Hänsel, Katrin; Cho, Sunghye; Berretta, Sarah; Mehta, Aarush; Nikzad, Amir; Guo, Jessica; **Cong, Yan**; Pradhan, Sameer; Fiumara, James; Liberman, Mark; Tang, Sunny. *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.

2019 Smith, Kaylin; Ma, Ye; **Cong, Yan**; Beretta, Alan. *The Psycholinguistics in Iceland— Parsing and Prediction (PiPP)*. Interactions between syntax and semantics: The semantic P600, University of Iceland. Reykjavík, Iceland.

2019 Smith, Kaylin; Ma, Ye; **Cong, Yan**; Beretta, Alan. *The 32nd CUNY Conference on Human Sentence Processing*. Semantic attraction in sentence processing. University of Colorado Boulder. Boulder, Colorado.

2019 **Cong, Yan**. *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.

2019 **Cong, Yan**. *International Workshop on Degrees and Grammar: An East Asian Perspective*. A subjective analysis of the Mandarin 'you' structure. Nanjing University, China.

2018 **Cong, Yan**. *The Annual Conference of the Chinese Language Teachers Association (CLTA)*. The non-canonical argument realization of the Chinese alternations. Washington, D.C.

2018 **Cong, Yan**. *The 12th Arizona Linguistics Circle (ALC-12)*. Rethinking the implicit argument in Kiswahili. University of Arizona. Tucson, Arizona.

2018 **Cong, Yan**. *The 2nd Crete Linguistics Summer School (CreteLing-2)*. A preliminary semantic analysis of predicates of taste in Chinese. University of Crete. Greece.

2017   Parrish, Alicia; Kelley, Patrick; Smith, Kaylin; **Cong, Yan**; Beretta, Alan. *Annual Meeting of The Society for the Neurobiology of Language (SNL)*. Prosodic lengthening and boundary prediction in nominal compounds: An ERP study. Baltimore, Maryland.

2017   **Cong, Yan**. *The Linguistic Society of America Linguistics Institute (LSA)*. T1 and T0 Sandhi in Tongliao Mandarin: an OT approach. University of Kentucky. Lexington, Kentucky.

2016   **Cong, Yan**. *The 28th North American Conference on Chinese Linguistics (NACCL-28)*. Markedness theory and the acquisition of the Mandarin potential complement construction. Brigham Young University. Salt Lake City, Utah.

Teaching and Administrative Experience

08/2023 – present   Assistant Professor, Purdue University, School of Languages and Cultures
- **Course: Statistical Modeling in Linguistics (Course No. LING598-LC596)**
- **Course: ornerstone: Critical thinking and communication (Course No. SCLA102)**

08/2017 – 12/2020   Teaching Assistant, Michigan State University, Department of Linguistics and Languages
- **Course: Introduction to Linguistics (Course No. LIN401, 7 semesters)**
- **Content**: Essentials of linguistics and applied linguistics
- **Task**: Gave (guest) lectures, organized in-class discussions and recitations, created and graded problem sets and exams, maintained online learning platforms
- **Format**: 2-hour lecture, 1-hour tutorial and recitation, 2 to 3 midterm exams. Usually face-to-face. Synchronous online during COVID
- **Students**: 90% U.S. students, with majors/minors ranging from TESOL, elementary education, computer science, to linguistics, speech and language pathology. Cap: 80.

08/2016 – 05/2017   Instructor, Michigan State University, Department of Linguistics and Languages
- **Course: Chinese language (Courses No. CHS101 and CHS102 1 semester; CHS411 1 semester)**
- **Contnt:** Elementary spoken Chinese, advanced spoken Chinese, Chinese writing/composition
- **Task**: Gave lectures, organized in-class discussions, created exams, graded homeworks
- **Format**: Twice a week, about 50 minutes each session (interactional teaching, classroom activities), 2 to 3 midterm exams. Face-to-face, combined with online learning platform and regular offline office hours.
- **Students**: 70% English speakers, 30% Heritage (Asian language) speakers, Cap: 60.

04/2015- 07/2016   Project Assistant, *Joint Research Center on Chinese Linguistics*, Hong Kong Polytechnic University, Department of Chinese and Bilingual Studies
- Worked directly with project managers to help implement project goals
- Assigned duties to staff and liaised with projects leads, as needed
- Oversaw variable aspects of Chinese linguistics projects, including grants arrangement

01 – 04/2015	Teaching Assistant, Hong Kong Baptist University, English Department

○ **Course: Sounds of English around the World; Pragmatics in Language Learning and Teaching (Courses No. ENGL2016 1 semester; LANG7630 1 semester)**
○ **Task**: Recitations, explained complex concepts to non-specialists, online teaching
○ **Students**: 70% Cantonese speakers; 30% Mandarin speakers, pre-service and in-service language instructors, future English teachers. Cap: 60.

05 - 08/2013	Intern, Mandarin Garden, Changchun, China

○ **Content**: Intermediate Chinese
○ **Task**: Shadowed senior instructors' teaching, graded students' compositions
○ **Material**: Self-designed; New Practical Chinese Reader
○ **Students**: Korean 50%; Russian 20%; American 10%; Others 20%. 20+ students.

2012 – 2013	Chinese Teacher, Jilin Normal University International Students Center, China

○ **Content**: Basic — Advanced Chinese
○ **Task**: HSK (*a TOEFL-like Chinese language proficiency test*) preparation, characters learning, grammatical constructions, conversation practice
○ **Material**: Self-designed, Boya Chinese, Integrated Chinese
○ **Students**: Korean 80%, American 10%, Heritage speakers 10%. 30+ students.

2012 – 2013	English Teacher, Jilin Normal University Community Center, China

○ **Content**: Middle school and High school English, College Entrance Subject Exam
○ **Task**: English exams preparation, English grammar, reading and writing strategies
○ **Material**: Self-designed, The National College Entrance Examination preparation test-sets, English subject exam test-sets
○ **Students**: High school and middle school students. 15+ students.

Research Experience

Communicative Disorders; Corpus linguistics 07/2021-present Postdoctoral researcher, Feinstein Institutes for Medical Research, Northwell Health: **Understanding and digitalizing mental illness using Machine Learning (ML) and Natural Language Processing (NLP)**

○ Shadow clinicians and psychiatrists, help interview mental disorders patients
○ Develop distributional semantic models using diverse parameters: context type, context window, frequency weighting, similarity measure
○ Deploy pre-trained transformer language models (GPT-3, BERT) to identify biomarkers in psychosis language and speech. Deep learning toolkit: PyTorch, TensorFlow
○ Develop NLP pipelines and scalable classifiers, leveraging CoreNLP (Semgrex, Dependency parse), Penn Discourse Treebank, WordNet-3.0
○ Large-scale corpus processing and annotating, calculating corpus linguistics features such as type-token-ratio, age of acquisition, utterance length, frequency and prevalence

Computational linguistics 09/2020- 08/2021	Graduate student researcher, Department of Linguistics and Languages, MSU: **Perform error analysis, increase language models' transparency**
	○ Deployed Google Cloud virtual machine instances to conduct task-based inference, analyze neural language models' functionalities, and design assessment algorithms
	○ Developed Python programs and R programs to evaluate transformers' performance, designing metrics such as *accuracy, cross-entropy loss, HITs@K, relative rank*
	○ Identified areas for language models improvement

Phonetics and Phonology 08/2019- 07/2021	Lab member, *Timing, Attention, and Perception Lab*, Department of Psychology, MSU: **Speech perception in noise; confusion matrix; sonority scales**
	○ Developed R scripts for confusion matrices analysis of speech perception in noise: Multidimensional Scaling (MDS) using the cmdscale() function in R
	○ Implemented and plotted MDS (package *igraph*) as layout.mds in R
	○ Used MATLAB to manipulate data and generate confusion matrix
	○ Modeled correlation of rhythm variation and speech perception in noise using SPSS and R (packages: *lattice,ggplot,dplyr,tidyverse*), resulting in 1 manuscript

Language experiment 08/2016- 08/2021	Data analyst / Graduate student researcher, *Semantics & Syntax Lab*, Department of Linguistics and Languages, MSU: **Psycholinguistics, Behavioral experiments**
	○ Designed 3 complex elicitation tasks collecting understudied language data
	○ Designed 2 artificial language learning experiments and 1 lexical decision task performed in PsychToolbox (MATLAB), PsychoPy - 2 & 3 (online), IBEX farm
	○ Developed 4 web-based acceptability judgment surveys and 1 self-paced reading task performed in Amazon Mechanical Turk (AMT) and Prolific
	○ Analyzed speech/text data (repeated measures ANOVAs (GG corrected) and pairwise comparisons). These studies led to publications

Neuro-physiology 08/2016- 08/2021	Lab member, *Neurolinguistics Lab*, Department of Linguistics and Languages, MSU: **Neurophysiology equipment, auditory and visual paradigms**
	○ Led coordination of stimuli design and paradigm design (3-8 team members)
	○ Provided coordination of data collection efforts, Electroencephalography (EEG) measurement for 50+ participants, resulting in 3 conference presentations
	○ Trained and monitored junior lab members in testing procedures

Corpus linguistics 04/2015- 07/2016	Project assistant, *Joint Research Center on Chinese Linguistics*, Hong Kong Polytechnic University: **Language data corpus, ontology**
	○ Balanced Corpus, Web-based Corpus, and Inter-language Corpus
	○ Assisted annotation, classification, and statistical modeling for 2 ontology projects on World Chineses Variations and Chinese Linguistic KnowledgeNet

Technical Skills

2017-present	Programming Languages
	○ Design, implement and debug **Python** programs
	○ Object-centered design and implementation in C++
	○ Statistical testing, modeling, advanced graphics in **R** and MATLAB
	○ Basic familiarity with **bash** scripting, JavaScript, HTML, CSS, SQL

2017-present	Sample scripts

- Using Transformer Language Models (LM) to detect language and speech disturbances in mental disorders `https://github.com/yancong222/SSD-LM-STanglab`
- Natural Language Processing (NLP) and Machine Learning (ML): `https://github.com/yancong222/scriptscz`
- (Computational) Experiments in linguistic meaning and beyond: `https://github.com/yancong222/scripts`

2018-present	Software development and implementation

- Development Environments: Visual Studio, RStudio, Anaconda
- Productivity Applications: Git/GitHub
- Cloud Service: Google Cloud Platform (GCP), Microsoft Azure
- Acoustics software: OpenSmile (INTERSPEECH), Montreal Forced Aligner (MFA in kaldi), Audacity, Praat
- Psychology software: PsychoPy (Visual Paradigm), E-Prime

Service and Leadership

Reviewer	Computer Speech & Language	2023-present
Reviewer	Scientific Reports	2023-present
Reviewer	Health Informatics Journal	2023-present
Member	Modern Language Association	2023-present
Reviewer	The Association for Computational Linguistics	2022-present
Mentor	Speech and Language Pathology Lab, Zucker Hillside Hospital	08/2021-present
Reviewer	Association of Contemporary African Linguistics	2020-present
Co-Chair	Colloquium committee, MSU Linguistics Department	05/2018-05/2020
Member	Society for the Neurobiology of Language	2018-present
Member	The Chinese Language Teachers Association – USA (CLTA-USA)	2018-2019

Fundings, Grants and Award

2022	**Educational Advancement Award**, *Advancing Women in Science and Medicine (AWSM)* *Computational linguistics can predict psychiatric diagnoses and outcome.* Northwell health $5,000
2020	**Dissertation Completion Fellowship**, MSU Graduate School $10,000 (2 Students Per College)
06/2018	**Student Fellowship**, *North American Summer School on Logic, Language, and Information (NASSLLI)*, at Pittsburgh, Pennsylvania, Carnegie Mellon University $500 (Travel and Accommodation Full Support)
2017-2020	**Summer Support Fellowship**, MSU Department of Linguistics and Languages $4,500 (4 times in a roll)
2015	**Best Paper Award**, *The 29th Pacific Asia Conference on Language, Information and Computation*, The invertible construction in Chinese. Shanghai Jiaotong University, China

Non-credit Professional Development Programs

2021	**European Summer School in Logic, Language and Information (ESSLLI)**, (virtually) Free University of Bozen-Bolzano, Italy

2020 **North American Summer School for Logic, Language, and Information (NASSLLI)**,
Carnegie Mellon University

2019 **Linguistic Society of America (LSA) Linguitics Institute**
University of California, Davis

2019 **Certificate of College Teaching Institute (CCT)**, *MSU Graduate School*
The University Graduate Certification in College Teaching

2018 **The 2ⁿᵈ Crete Linguistics Summer School (CreteLing-2)**
University of Crete, Greece

2018 **Linguistic Society of America (LSA) Linguistics Institute**
University of Kentucky

Languages

Chinese	native level	普通话水平测试一级乙等	*Top level division B in Putonghua Proficiency Test*
English	near-native level		*MSU TA English proficiency test: High pass*
Cantonese,Korean	conversational level		*Limited working proficiency*
Kiswahili,Fijian	elementary level		*4 months of fieldwork elicitation*