

EDUCATION

Ph.D. in Computer Science

2023

Université de Lorraine

Nancy, France

- **Research topics:**

- * Discourse Analysis: Developed semi- and unsupervised methods for discourse structure extraction and relation prediction in dialogues.
- * Machine Learning for Health: Applied machine learning techniques to identify linguistic markers associated with mental illnesses (e.g., schizophrenia, depression).

- **Supervisors:** Maxime Amblard (LORIA), Chloé Braud (CNRS).

- **Research Internship:** University of British Columbia, supervised by Giuseppe Carenini (2021-2022).

M.Sc. in Computational Linguistics

2019

Université Paris Cité (formerly Paris Diderot)

Paris, France

- Specialized in Natural Language Processing (NLP), graduated *summa cum laude*.

Master in Strategy & Management

2017

ESSEC Business School

Paris, France

- Graduated *summa cum laude*.

B.A. in French

2015

Wuhan University of Technology

Wuhan, China

- Three-time recipient of National Scholarship (highest honor); graduated *summa cum laude*.

PROFESSIONAL EXPERIENCE

Post-doctoral Researcher

2023-present

Computer Science Department, University of British Columbia

Vancouver, Canada

- **Research projects:**

- * Large Language Models: In-context learning (ACL 2025); reinforcement learning (EMNLP 2025); multi-agent systems (collaboration with Salesforce AI Research).
- * Discourse Analysis: LLM-based discourse parsing (SIGDIAL 2024, **Best Paper Award**); benchmarking; interpretability of discourse relations.
- * Machine Learning for Healthcare: Dementia detection with multimodal data (PMLR 2025).

- **Supervisor:** Giuseppe Carenini, carenini@cs.ubc.ca

Temporary Teaching and Research Associate (ATER)

2022-2023

Institut des sciences du Digital, Management & Cognition, Université de Lorraine

Nancy, France

- 98 hours of teaching (plenary lectures and tutorials); courses: programming languages, introduction to NLP, formal logic and reasoning, discourse and dialogue.

Teaching Assistant during Ph.D.

2019-2022

LORIA, Inria, Université de Lorraine

Nancy, France

Industrial Experience

Synomia, Paris, France

2019.5-8

- Research Project: Developed an automatic language processing chain for Chinese for automatic extraction of entities and sentiments. The processing chain includes word segmentation, syntactic analysis, and entity extraction.

SemantiWeb France, Paris, France

2018.3-8

- Research Project: Machine learning for sentiment detection in SNCF-related tweets; linguistic resource development (segmentation, dictionaries, named entities).

Lionbridge, remote, in collaboration with Amazon Alexa

2018.5-6

- Research Project: Designed grammars for Alexa Chinese speech recognition using Backus–Naur Form (BNF).

Van Cleef & Arpels, Paris, France

2016-2017

- CRM Coordinator, Europe: Responsible for data analysis and database management.

REFERRED CONFERENCE/JOURNAL ARTICLES

(1) Topic-Guided Reinforcement Learning with LLMs for Enhancing Multi-Document Summarization.
Chuyuan Li, Austin Xu, Shafiq Joty, and Giuseppe Carenini.

In *Findings of the Association for Computational Linguistics: EMNLP 2025*, Suzhou, China. Association for Computational Linguistics. 2025.

Link: <https://arxiv.org/pdf/2509.09852>

(2) CEMTM: Contextual Embedding-based Multimodal Topic Modeling.

Amirhossein Abaskohi, Raymond Li, **Chuyuan Li**, Shafiq Joty, and Giuseppe Carenini

In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Suzhou, China. Association for Computational Linguistics. 2025.

Link: <https://arxiv.org/pdf/2509.11465>

(3) Delta-KNN: Improving Demonstration Selection in In-Context Learning for Alzheimer’s Disease Detection.

Chuyuan Li, Raymond Li, Thalia S. Field, and Giuseppe Carenini

In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 25807-25826, Vienna, Austria. Association for Computational Linguistics. 2025.

Link: <https://aclanthology.org/2025.acl-long.1253/>

(4) Explicit Bayesian Inference to Uncover the Latent Themes of Large Language Models.

Raymond Li, **Chuyuan Li**, Gabriel Murray, and Giuseppe Carenini

In *Findings of the Association for Computational Linguistics (ACL): ACL 2025*, pages 21819-21833, Vienna, Austria. Association for Computational Linguistics. 2025.

Link: <https://aclanthology.org/2025.findings-acl.1123/>

(5) Multimodal Classification of Alzheimer’s Disease by Combining Facial and Eye-tracking Data.

Shih-Han Chou, Miini Teng, Harshinee Sriram, **Chuyuan Li**, Giuseppe Carenini, Cristina Conati, Thalia S. Field, Hyeju Jang, Gabriel Murray

In *Proceedings of the 4th Machine Learning for Health Symposium (ML4Health)*, pages 219–232, Vancouver, Canada. PMLR. 2025.

Link: <https://proceedings.mlr.press/v259/chou25a.html>

(6) Dialogue Discourse Parsing as Generation: a Sequence-to-Sequence LLM-based Approach.

Chuyuan Li, Yuwei Yin, and Giuseppe Carenini.

In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGdial)*, pages 1-14, Kyoto, Japan. Association for Computational Linguistics. 2024. **Best Paper Award.**

Link: <https://aclanthology.org/2024.sigdial-1.1/>

(7) Coherence-based Dialogue Discourse Structure Extraction using Open-Source Large Language Models.

Gaetano Cimino, **Chuyuan Li**, Giuseppe Carenini, and Vincenzo Deufemia

In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGdial)*, pages 297-316, Kyoto, Japan. Association for Computational Linguistics. 2024.

Link: <https://aclanthology.org/2024.sigdial-1.26/>

(8) Discourse Structure Extraction from Pre-Trained and Fine-Tuned Language Models in Dialogues.

Chuyuan Li, Patrick Huber, Wen Xiao, Maxime Amblard, Chloé Braud and Giuseppe Carenini.

In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 2562–2579, Dubrovnik, Croatia. Association for Computational Linguistics. 2023.

Link: <https://aclanthology.org/2023.findings-eacl.194/>

(9) Multi-Task Learning for Depression Detection in Dialogs.

Chuyuan Li, Chloé Braud and Maxime Amblard.

In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGdial)*, pages 68–75, Edinburgh, UK. Association for Computational Linguistics. 2022.

Link: <https://aclanthology.org/2022.sigdial-1.7/>

(10) Investigation par méthodes d’apprentissage des spécificités langagières propres aux personnes avec schizophrénie (Investigating Learning Methods Applied to Language Specificity of Persons with Schizophrenia).

Maxime Amblard, Chloé Braud, **Chuyuan Li**, Michel Musiol, Caroline Demily et Nicolas Franck.
In *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)*. Volume 2: Traitement Automatique des Langues Naturelles. 2020.
Link: <https://www.aclweb.org/anthology/2020.jeptalnrecital-taln.2/>

(11) Characterizing the Response Space of Questions: data and theory.

Jonathan Ginzburg, Zulipiye Yusupujang, **Chuyuan Li**, Kexin Ren, Aleksandra Kucharska and Pawel Lupkowski.
In *Journal Dialogue & Discourse (D&D)* 13.2 (2022): 79-132.
DOI: <https://doi.org/10.5210/dad.2022.203>

(12) Characterizing the Response Space of Questions: a Corpus Study for English and Polish.

Jonathan Ginzburg, Zulipiye Yusupujang, **Chuyuan Li**, Kexin Ren and Pawel Lupkowski.
In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue (SIGdial)*, pages 320–330, Stockholm, Sweden. Association for Computational Linguistics. 2019.
Link: <https://www.aclweb.org/anthology/W19-5937/>

ARTICLES IN PEER-REVIEWED WORKSHOP PROCEEDINGS

(13) Discourse Relation Recognition with Language Models Under Different Data Availability

Shuhaib Mehri, **Chuyuan Li**, Giuseppe Carenini
In *Proceedings of the 6th Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences (CODI 2025)*, EMNLP, pages 148–156, Suzhou, China. Association for Computational Linguistics. 2025. *To appear*.

(14) The DISRPT 2025 Shared Task on Elementary Discourse Unit Segmentation, Connective Detection, and Relation Classification

Chloé Braud, Amir Zeldes, **Chuyuan Li**, Yang Janet Liu, Philippe Muller
In *Proceedings of the 4th Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2025)*, EMNLP, pages 66–85, Suzhou, China. Association for Computational Linguistics. 2025. *To appear*.

(15) Hierarchical Attention Adapter for Abstractive Dialogue Summarization

Raymond Li, **Chuyuan Li**, Gabriel Murray, Giuseppe Carenini
In *Proceedings of the 5th New Frontiers in Summarization Workshop (NewSumm)*, EMNLP, Suzhou, China. Association for Computational Linguistics. 2025. *To appear*.

(16) Multi2: Multi-Agent Test-Time Scalable Framework for Multi-Document Processing

Juntai Cao, Xiang Zhang, Raymond Li, Jiaqi Wei, **Chuyuan Li**, Shafiq Joty, Giuseppe Carenini
In *Proceedings of the 5th New Frontiers in Summarization Workshop (NewSumm)*, EMNLP, Suzhou, China. Association for Computational Linguistics. 2025. *To appear*.
Link: <https://arxiv.org/pdf/2502.20592>

(17) On Large Foundation Models and Alzheimer's Disease Detection

Chuyuan Li, Giuseppe Carenini, and Thalia Field
In *Proceedings of the Second Workshop on Patient-Oriented Language Processing (CL4Health)*, NAACL, pages 158–168, Albuquerque, New Mexico. Association for Computational Linguistics. 2025.
Link: <https://aclanthology.org/2025.cl4health-1.13/>

(18) Discourse Relation Prediction and Discourse Parsing in Dialogues with Minimal Supervision

Chuyuan Li, Chloé Braud, Maxime Amblard, and Giuseppe Carenini
In *Proceedings of the 5th Workshop on Computational Approaches to Discourse (CODI)*, pages 161-176, St. Julians, Malta. Association for Computational Linguistics. 2024.
Link: <https://aclanthology.org/2024.codi-1.15>

(19) Investigating non lexical markers of the language of schizophrenia in spontaneous conversations.

Chuyuan Li, Maxime Amblard, Chloé Braud, Caroline Demily, Nicolas Franck and Michel Musiol.
In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse (CODI)*, pages 20–28, Punta Cana, Dominican Republic and Online. Association for Computational Linguistics. 2021.
Link: <https://aclanthology.org/2021.codi-main.3/>

(20) The Annotators Did Not Agree on Some of the Guidelines Examples.

Maxime Amblard, Karën Fort, Bruno Guillaume, Philippe de Groote, **Chuyuan Li**, Pierre Ludmann, Michel Musiol, Siyana Pavlova, Guy Perrier and Sylvain Pogodalla.
In *Proceedings of the 17th Joint ACL - ISO Workshop on Interoperable Semantic Annotation (ISA-17)*, 2021.

Link: <https://sigsem.uvt.nl/isa17/AmblardEtAl-final.pdf>

OTHERS

(21) A Semi-supervised Dialogue Discourse Parsing Pipeline.

Chuyuan Li, Maxime Amblard and Chloé Braud.

In *Actes des 5èmes journées du Groupement de Recherche CNRS “Linguistique Informatique, Formelle et de Terrain”*. (2023): 85.

Link: <https://hal.science/hal-04313917/document#page=93>

(22) Facing Data Scarcity in Dialogues for Discourse Structure Discovery and Prediction.

Chuyuan Li.

Doctoral dissertation, Université de Lorraine, 2023.

Link: https://theses.hal.science/tel-04218515/file/DDOC_T_2023_0107_LI.pdf

AWARDS & HONORS

2024	Best Paper Award , 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue
2024	Postdoctoral Fellows Travel Award (\$1,500) , UBC Faculty of Science
2022	DrEAM International Mobility Grant (€9,000) , Université de Lorraine & Mitacs
2019	Summa Cum Laude , M. Sc., Université Paris Cité, France
2017	Summa Cum Laude , Master, ESSEC Business School, France
2015	Summa Cum Laude , Bachelor, 1st in class, Wuhan University of Technology, China
2011-2014	National Scholarship (¥8,000/year) , Wuhan University of Technology, China

SERVICE

Co-Organizer	SHROOM-CAP 2025, “Shared Task on Hallucinations and Related Observable Overgeneration Mistakes”, held at AACL DISRPT 2025, “Shared Task on Discourse Segmentation, Connective and Relation Identification across Formalisms”, held at EMNLP CODI-CRAC 2025, “First Joint Workshop on Computational Approaches to Discourse and Computational Models of Reference, Anaphora and Coreference”, held at EMNLP CODI 2024, “Workshop on Computational Approaches to Discourse” IWCS 2023, “The 15th International Conference on Computational Semantics” JEP-TALN-RECITAL 2020, “The 27th French National Joint Conference”
Area Chair	Conference of the European Chapter of the ACL (EACL), 2024, 2026 Conference of the Nations of the Americas Chapter of the ACL (NAACL), 2025 International Conference on Computational Linguistics (COLING), 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2024
Program Committee	International Conference on Language Resources and Evaluation (LREC), 2026 Annual AAAI Conference on Artificial Intelligence (AAAI), 2025 Conference on Computational Semantics (IWCS), 2025 Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGdial) 2024, 2025
Reviewer	ACL Rolling Review (standing reviewer since 2023) Transactions of the Association for Computational Linguistics (TACL), 2024 International Conference on Computational Linguistics (COLING), 2020-2024

SUPERVISION

Ph.D. student	Gaetano Cimino, University of Salerno, Sep. 2023 – July 2024
Master student	Albert Millert et Anar Yeginbergenova, University of Lorraine NLP, Sep. 2020 – Sep. 2021
Undergraduate	Armin Talaie, UBC Computer science, Oct. 2024 – July 2025 Ahmad Shuhaib Mehri, UBC Computer science, Sep. 2023 – Aug. 2024 Ryan Lazenby, UBC Computer science, Dec. 2023 – Apr. 2024

LANGUAGES & SKILLS

Natural Language (professional proficiency): Chinese, French, English

Programming Language: Python, Bash

Machine Learning Toolkit: Huggingface, PyTorch, TensorFlow, AllenNLP, Scikit-Learn

Operating system: Mac OS, Linux