Chuyuan Li

Department of Computer Science, The University of British Columbia

web: chuyuanli.github.io
email: chuyuan.li@ubc.ca

PROFESSIONAL SUMMARY

My research focuses on natural language processing (NLP), particularly on developing novel algorithms and systems that advance Large Language Models' capabilities in language understanding, reasoning, and generation. My background combines computational linguistics and a Ph.D. in computer science, strengthened by international research collaborations across Europe and North America.

EDUCATION

Ph.D. in Computer Science
2023

Lorraine Research Laboratory in Computer Science and its Applications & Université de Lorraine
France

- **Dissertation:** Facing Data Scarcity in Dialogues for Discourse Structure Discovery and Prediction.
- **Research Internship:** University of British Columbia, supervised by Giuseppe Carenini (2021-2022).

M.Sc. in Computational Linguistics
2019

Université Paris Cité (formerly Paris Diderot)
France

- Specialized in natural language processing; graduated with distinction (*mention très bien*).

B.A. in French
2015

Wuhan University of Technology
China

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

PROFESSIONAL EXPERIENCE

Post-doctoral Researcher
Nov 2023-Present

Computer Science Department, University of British Columbia
Canada

- **Research projects:**
  * Large Language Models: In-context learning (ACL 2025); reinforcement learning (EMNLP 2025); test-time scaling (NewSumm workshop 2025); multi-agent systems (collaboration with Salesforce AI Research).
  * Discourse Analysis: LLM-based discourse parsing (SIGDIAL 2024, **Best Paper Award**); relation prediction (CODI workshop 2024-25); benchmarking (BeDiscovER, under submission), and interpretability.
  * Machine Learning for Health: Dementia detection with multimodal data (PMLR 2025); principled zero-shot (CL4Health workshop 2025) and in-context learning approaches (ACL 2025).
- **Supervisor:** Giuseppe Carenini, carenini@cs.ubc.ca

Temporary Teaching and Research Associate (half-time ATER)
Sep 2022-Sep 2023

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

- 100 hours of teaching, including plenary lectures and tutorials to L1-M2 students.
- Subjects: introduction to NLP, formal logic and reasoning, discourse and dialogue, introduction to Grid'5000 HPC, programming languages (Python, C).

Teaching Assistant during Ph.D.
Sep 2019-Sep 2022

LORIA, Inria, Université de Lorraine
France

Industrial Experiences

***Synomia**, Paris, France*
May-Sep 2019

- Computational linguist: 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*
Jul-Aug 2018

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

Lionbridge**, remote, in collaboration with **Amazon Alexa
May-Jun 2018

- Computational linguist: Created and implemented Backus–Naur Form (BNF) grammars to improve Chinese speech recognition for Alexa.

PUBLICATIONS

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. 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. 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. 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. 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. 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. 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. 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. 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 Yusupujiang, **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 Yusupujiang, **Chuyuan Li**, Kexin Ren and Paweł Łupkowski.
In *Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue (SIGdial)*, pages 320–330, Stockholm, Sweden. 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. 2025.
Link: https://aclanthology.org/2025.codi-1.13/

(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. 2025.
Link: https://aclanthology.org/2025.disrpt-1.1/

(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*, pages 17–30, Hybrid. 2025.
Link: https://aclanthology.org/2025.newsum-main.2/

(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*, pages 135–156, Hybrid. 2025.
Link: https://aclanthology.org/2025.newsum-main.10/

(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. 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. 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. 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) BeDiscovER: The Benchmark of Discourse Understanding in the Era of Reasoning Language Models.
Chuyuan Li and Giuseppe Carenini.
In *ACL Rolling Review Preprint.* October 2025. In submission to *ACL 2026.
Link: https://arxiv.org/pdf/2511.13095

(22) 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

(23) 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** (3 years), Ministry of Education of China

Service

Co-Organizer	CODI-CRAC'26, "2nd Joint Workshop on CODI (7th) and CRAC (9th)", to be held at ACL 2026
	DISRPT'25, "Shared Task on Discourse Segmentation, Connective and Relation Identification across Formalisms", held at CODI-CRAC'25
	CODI-CRAC'25, "1st Joint Workshop on CODI (6th) and CRAC (8th), EMNLP 2025
	CODI'24, "5th Workshop on Computational Approaches to Discourse, Context and Document-Level Inferences", EACL 2024
	IWCS 2023, "15th International Conference on Computational Semantics"
	JEP-TALN-RECITAL 2020, "National Joint Conference on Speech Studies (33rd), Natural Language Processing (27th), and Student Researchers in Computer Science for Natural Language Processing (22nd)", France
Area Chair	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
	Conference of the European Chapter of the ACL (EACL), 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
	Workshop on Linguistic Analysis for Health (HeaLing), 2026
	Workshop on Patient-Oriented Language Processing (CL4Health), 2025-26
Reviewer	ACL Rolling Review (standing reviewer since 2023)
	Transactions of the Association for Computational Linguistics (TACL), temporal 2024

Supervision

Ph.D. student	Raymond Li (research mentorship), UBC, since 2025
	Gaetano Cimino, University of Salerno, Sep 2023-Jul 2024 $\Rightarrow$ Now postdoc at UBC
Master student	Amirhossein Abaskohi, UBC, since 2025 $\Rightarrow$ Now Ph.D. student at UBC
	Albert Millert, University of Lorraine, Sep 2020-Sep 2021
	Anar Yeginbergenova, University of Lorraine, Sep 2020-Sep 2021
Undergraduate	Armin Talaie, UBC, Oct 2024-Jul 2025
	Ahmad Shuhaib Mehri, UBC, Sep 2023-Aug 2024 $\Rightarrow$ Now Ph.D. student at UIUC
	Ryan Lazenby, UBC, Jan 2024-Jun 2024

Languages & Skills

Languages: Chinese (native), English (fluent), French (fluent)
Programming Languages: Python, Bash
Machine Learning: PyTorch, TensorFlow, HuggingFace, DeepSpeed, Weights & Biases, Scikit-Learn
High-Performance Computing: Grid'5000 (OAR), Compute Canada (Slurm)
Operating Systems: macOS, Linux