

BriGap 2026

**Third Workshop on the Bridges and Gaps between Formal
and Computational Linguistics**

Proceedings of the Workshop

July 11, 2026

The BriGap organizers gratefully acknowledge the support from the following sponsors.

Founded by



EFL



©2026 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN

Introduction

We are excited to welcome you to BriGap-3 in Paris, France!

This third edition of the workshop on Bridges and Gaps between Formal and Computational Linguistics is the first to be organised independently of any other conference. It follows up on the first edition in 2022 (co-located with ESSLLI) and the second in 2025 (co-located with IWCS). We have decided to continue with the dual submission policy implemented for the second edition, soliciting archival papers to be published in the ACL Anthology, and non-archival submissions describing work in progress or work published elsewhere that would be of interest to our audience.

We are very pleased with the response to the call for this third edition. Out of the submissions received, 21 were selected for presentation (8 oral presentations and 13 poster presentations), of which 13 are included in these proceedings. These contributions sit at the intersection of theoretical linguistics and computational linguistics or natural language processing, spanning a wide range of topics such as formal semantics and inference, language models and linguistic theory, pragmatics and common ground, grammar engineering and low-resource languages, and human–model comparison.

In addition to these contributed presentations, BriGap-3 includes two invited talks. Raquel Fernández (Universiteit van Amsterdam) will explore how machine learning models trained on body movements, vision, and language can shed light on multimodality in human cognition, whereas Tal Linzen (New York University) will show how formal symbolic models can be leveraged to make LLM pretraining more efficient, evaluate their capabilities, and improve how they update their beliefs during interaction. Together, the two keynotes bridge cognitively and formally grounded approaches to language, setting the stage for a rich exchange between formal and computational linguistics.

BriGap-3 was made possible thanks to the financial support of RT LIFT2, a France-based research group aiming to bring together researchers in computational linguistics, formal linguistics, and field linguistics around shared questions, data, and tools, LLF, a joint Université Paris Cité-CNRS laboratory doing research covering a wide range of topics in formal, experimental and computational linguistics, and InIdEx EFL, an Université Paris Cité project that supports innovative research on languages and language, bringing together different fields of linguistics and related disciplines (psychology, neuroscience, computer science) around themes that respond to the major challenges of today’s world.

Timothée Bernard, Giulia Rambelli, Emmanuele Chersoni, Program Chairs

Program Committee

Senior Program Committee

Timothée Bernard, Université Paris Cité, CNRS, Laboratoire de linguistique formelle
Emmanuele Chersoni, Hong Kong Polytechnic University
Giulia Rambelli, University of Aix-Marseille, CNRS, Laboratoire Parole et Langage / Laboratoire d'Informatique et Systèmes

Program Committee

Lasha Abzianidze, Utrecht University
Pascal Amsili, Sorbonne Nouvelle (Paris 3)
Giosuè Baggio, Norwegian University of Science and Technology
Sacha Beniamine, University of Surrey
Philippe Blache, Aix-Marseille University, CNRS
Johan Bos, University of Groningen
Chloe Braud, IRIT, CNRS, ANITI
Canaan Breiss, University of Chicago
Marie Candito, Université Paris Cité
Yan Cong, Purdue University
Benoît Crabbé, Université Paris Cité
Luca Dini, Institute of Computational Linguistics (CNR, Pisa)
Abdellah Fourtassi, Aix-Marseille University
Yu-Yin Hsu, Hong Kong Polytechnic University
Gene Kim, University of South Florida
Dan Lassiter, University of Edinburgh
Timothée Mickus, University of Helsinki
Koji Mineshima, Keio University
Filip Miletic, University of Stuttgart
Ludovica Pannitto, University of Bologna
Chiara Paolini, KU Leuven
Denis Paperno, Utrecht University
Vered Shwartz, University of British Columbia
Ece Takmaz, Utrecht University
Kees Van Deemter, Utrecht University
Grégoire Winterstein, Université du Québec à Montréal
Guillaume Wisniewski, Université Paris Cité
Olga Zamaraeva, University of Helsinki
He Zhou, Hong Kong Polytechnic University

Keynote Talk
Modelling Multimodality in Human Cognition with Machine Learning Models

Raquel Fernández
Universiteit van Amsterdam

Abstract: Linguistic communication is both inherently multimodal and interactive. Conceptual knowledge encoded in language is grounded in our sensory-motor experience, while in face-to-face dialogue we use speech in tandem with non-verbal signals such as gaze and gestures. In this talk, I will present our work on using machine learning models trained on body movements, vision, and language to study fundamental questions regarding multimodality in human cognition. I will first focus on behaviour: I will discuss our approach to co-speech gesture representation learning, showing that the resulting gesture embeddings exhibit properties that support theoretically motivated hypotheses. I will then move to describing our experiments on modelling brain activity with pre-trained vision-language models, which add to existing evidence for the intricate relationship between language and perception in the human brain.

Keynote Talk
Formal symbolic models for LLMs: pretraining, evaluation and post-training

Tal Linzen
New York University

Abstract: Formal symbolic models—ideal versions of the computational problems involved in learning about and interacting with the world—are central in linguistics and cognitive science. These models make it possible to generate unlimited amounts of synthetic data of variable complexity, as well as define verifiably correct outcomes for each instance of the problem. I will discuss three studies that leverage these properties of formal models. First, I will show that by pretraining transformer LLMs on formal languages before training them on natural languages, we can make training more compute and data efficient overall. Second, I will introduce context-free language recognition as an evaluation task for LLM. I will show that the complexity of the grammar reliably predicts the model’s accuracy on this task, and that even the strongest reasoning models available struggle to perform this task as the complexity of the language increases. Finally, I will demonstrate how symbolic Bayesian models can be used to evaluate and improve LLM abilities to update their probabilistic beliefs when interacting with users.

Table of Contents

<i>From Execution to Exploration: Bridging the Usability Gap in Formal Natural Language Inference</i> Koharu Saeki and Daisuke Bekki	1
<i>Neural Wani: Toward Accelerating the Automated Theorem Prover wani for Dependent Type Theory</i> Nanako Miyagawa, Hinari Daido and Daisuke Bekki	12
<i>Cross-linguistic Geometry of Adjective Representations in Multilingual Transformers: Semantic Class, Gradability, and Positional Effects.</i> Tancredi Monterosso	22
<i>Diagnosing Compositional Generalization in Transformers on ReCOGS with Compositional Graph Similarity</i> Bruno Franco and Edson Scalabrin	31
<i>Polar Questions in SPA–TTR: Linking Dialogue, Acquisition, and Neurosemantics</i> Jonathan Ginzburg, Shiyun Dong, Robin Cooper, Andy Lücking and Staffan Larsson	40
<i>Artificial Language Learning Paradigm Reveals Pragmatic Blind Spots in Vision-Language Models</i> Yan Cong and Julia Rayz	50
<i>Using the Mimi codec for metalinguistic representations</i> Artem Saloev, Erin Pacquetet and Nicolas Ballier	63
<i>Misalignments in Common Ground as a Bridge Between Pragmatic Theory and LLM Evaluation</i> Judith Sieker and Sina Zarriß	74
<i>Towards Benchmarking Old Church Slavonic Lemmatization</i> Usman Nawaz, Marianna Napolitano, Iris Karafillidis, Liliana Lo Presti and Marco Cascia ...	82
<i>Transformers Learning Contrafactuals: The Importance of Data Distributions</i> David Strohmaier and Simon Wimmer	94
<i>Grammar Engineering Meets LLMs: Development of Cantonese and Irish ParGram Treebanks</i> Chit-Fung Lam and Elaine Uí Dhonnchadha	121
<i>A Formal Model of Lexical Negation in Discrete Communication</i> Mikołaj Piotr Golecki and Timothée Bernard	135
<i>A graph-based analysis of semantic types and coercion in contextualized word embeddings</i> Long Chen and Deniz Yavas	148