BriGap 2025

The Second Workshop on the Bridges and Gaps between Formal and Computational Linguistics

Proceedings of the Workshop

September 24, 2025

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

Funded by

Order copies of this and other ACL proceedings from:

Introduction

We are excited to welcome you to BriGap-2, co-located with IWCS 2025, in Düsseldorf, Germany!

This second edition of the workshop on Bridges and Gaps between Formal and Computational Linguistics follows up on the first edition in 2022. We have implemented a number of changes that we hope will reflect the diverse communities that we aim to bring together in this event. In particular, we have worked towards designing an inclusive and welcoming submission policy, soliciting archival papers to be published in the ACL Anthology, non-archival abstracts describing work in progress, as well as presentations of already published articles that would be of interest to our audience.

We are especially proud of the success encountered in this second edition, with 11 publications presented at the workshop, 9 of which are included in these proceedings. The works range across a number of topics, including but not limited to Dependent Type Semantics, the syntactic abilities of LLMs, Lexical Functional Grammer, as well as the use NLP systems for cognitive science. We hope that future editions of the workshop will be able to build upon this success and continue to foster the diversity of topics addressed.

Beyond these 11 presentations, the workshop also includes two invited talks. Anna Rogers (ITU Copenhagen) will discuss data contamination in the age of LLMs, whereas Kees van Deemter (University of Utrecht) will address hallucinations and how to classify them. Both keynotes provide valuable perspectives on pitfalls and caveats of modern NLP technology, and provide an excellent starting point for a broader discussion on how to build successful interactions between formal and computational linguistics.

The BriGap-2 workshop 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.

We also want to thank our colleague Grégoire Winterstein, who helped us put together the workshop proposal before withdrawing from the organizing committee due to an excessive workload.

Timothée Bernard, Timothee Mickus, Program Chairs

Organizing Committee

Program Chairs

Timothée Bernard, Université Paris Cité, CNRS, Laboratoire de linguistique formelle
Timothee Mickus, University of Helsinki

Program Committee

Program Chairs

Timothée Bernard, Université Paris Cité
Timothee Mickus, University of Helsinki

Reviewers

Lasha Abzianidze, Pascal Amsili, Carolyn Jane Anderson

Olivier Bonami, Chloé Braud, Canaan Breiss, Tommi Buder-Gröndahl

Maria Copot, Benoit Crabbé

Isabelle Dautriche, Marie-Catherine De Marneffe

Katrin Erk

Richard Futrell

Juan Luis Gastaldi

Aurelie Herbelot

Gene Louis Kim

Dan Lassiter, Alessandro Lenci, Tal Linzen

Koji Mineshima, Teruyuki Mizuno

Denis Paperno, Christopher Potts

Christian Retoré

Kees van Deemter

Guillaume Wisniewski

Olga Zamaraeva

Keynote Talk
Studying Generalization in the Age of Contamination

Anna Rogers
IT University of Copenhagen
2025-09-24 09:30:00 – Room: **Room 3**

Abstract: In the age of Large Language Models, we can no longer be sure that the test data was not observed in training. This talk discusses the main approaches to studying generalization, and presents a new framework for working with controlled test-train splits across linguistically annotated data at scale.

Bio: Anna Rogers is an tenured Associate Professor at the IT University of Copenhagen is one of the foremost experts in Natural Language Processing (NLP). Her expertise ranges from ethics in NLP to frame semantics, and from computational social science to interpretability. Her contribution to the field goes beyond widely acclaimed scientific articles; she has also taken on significant responsibilities within the community, including heading the scientific committee of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023), and more recently taking on the role of co-editor-in-chief for the ACL Rolling Review platform.

Her recent work at the IT University of Copenhagen focuses on understanding large language models from a sociotechnical perspective. This has led her to studying the impact of data on what these models converge to, as well as how to make NLP models more efficient, transparent and reliable.

Keynote Talk
Classifying Hallucinations in Data-Text NLG: Avoiding the Pitfalls

Kees van Demter
Universiteit Utrecht
2025-09-24 14:20:00 – Room: **Room 3**

Abstract: Algorithms that produce textual output can sometimes "hallucinate", producing texts that express information that differs from what is required. In this presentation, I will talk about hallucination in Data-Text NLG, focusing on situations in which the task of the algorithm is to express a known body of information both fully and accurately. Various attempts have been made to clarify the notion of hallucination, and to distinguish between different types of hallucinations that can occur in the above-mentioned situations. I will examine some of these classifications and ask:
(1) Are the existing classifications well defined? (2) How feasible in practice is it to apply these classifications to concrete cases of Data-Text NLG? (This is joint work with Eduardo Calo and Albert Gatt, both at Utrecht University.) (3) How useful are the distinctions that these classifications make, for example for determining the seriousness of a hallucination, or for redesigning the NLG algorithm so as to avoid hallucinations? And finally, if time permits (4) What does our investigation tell us about hallucinations in other NLG situations, for instance in Question-Answering?

Bio: Kees van Deemter is an Emeritus Professor at the University of Utrecht, where he has been a major support and proponent of research in computational linguistics since 2018. As a long-standing expert in the area of Natural Language Generation, he focuses on structured inputs and their limits: his work ranges from logic-to-text systems to vagueness in natural language, and from referring expressions to some of the caveats of modern neural NLP systems.
Luckily for us, he has recently written an "autoworkography", so we can point you to his own words, which without a doubt will do a much better job of retracing his steps than we can: https://arxiv.org/abs/2504.04142.

Table of Contents