

Preface

François Coste

Univ Rennes, Inria, CNRS, IRISA, Rennes, France

FRANCOIS.COSTE@INRIA.FR

Faissal Ouardi

Faculty of Sciences, Mohammed V University in Rabat, Morocco

F.OUARDI@UM5R.AC.MA

Guillaume Rabusseau

University of Montreal, DIRO / Mila, CIFAR AI Chair, Canada

GRABUS@IRO.UMONTREAL.CA

These proceedings contain the works that were presented during the 16th edition of the International Conference on Grammatical Inference (ICGI). The conference was hosted by the Faculty of Sciences of Mohammed V University in Rabat, Morocco from July 10th to July 13th, 2023. It is the first time that the conference was held on the African continent.

Out of the 17 regular paper submissions reviewed by the ICGI Program Committee, 13 were accepted and included in the conference program. In addition, 2 extended abstracts on *work in progress* were accepted (after a lighter review process from the conference chairs ensuring relevance of scope and clarity) and presented during the conference. This edition of ICGI was partner with the *Transformers+RNN: Algorithms to Yield Simple and Interpretable Representations* (TAYSIR) competition (<https://remieyraud.github.io/TAYSIR/>), an online challenge on extracting simpler models from already trained neural networks. The conference included a special session presenting the competition with an opportunity for competitors to present their approach; a paper introducing the competition (authored by the competition’s organizers) as well as two short papers from participants presenting their methods are included in these proceedings.

This edition of the conference marked its 30 years anniversary. To celebrate this milestone, one of the founders of the Grammatical Inference field, Dana Angluin, was invited to give a distinguished lecture titled “*Learning of Regular Languages by Recurrent Neural Networks?*” and the *ICGI Angluin Tutorial Session* was inaugurated. This recurring session for future ICGI editions will consist in inviting an early career researcher to give a tutorial on the first day of ICGI. This year, Will Merrill was the first researcher to be invited to give this tutorial, on the topic of *Formal languages and neural models for learning on sequences*. It is worth mentioning that, even though it was not yet titled as the *ICGI Angluin Tutorial Session*, Gail Weiss’s tutorial on *Formal Abstractions of Neural Sequence Models* at ICGI 2020/21 served as a precursor. In addition to the distinguished lecture and tutorial, which took place on the first day of the conference, two invited speakers gave keynote talks during the conference: Cyril Allauzen (Google, New York) gave a presentation titled *Weighted Finite Automata (WFA) with Failure Transitions: Algorithms and Applications*, and Ahmed Elnaggar (Technische Universität München, Germany) gave a presentation titled *A journey into the Generative AI and large language models: From NLP to BioInformatics*. These 4 invited talks offered a broad and diverse view of Grammatical Inference, navigating from its roots strongly tied with the early ages of neural networks, to the applications of WFA with failure transitions to language modeling and of large language models to biological sequences.

These proceedings contain a diverse set of contributions in the field of Grammatical Inference. Several papers tackled the challenging setting of active learning, ranging from practical algorithms to learn finite state machines from data streams, to online change point detection and deriving lower bounds for active learning of automata. Aside the *Angluin tutorial* and TAYSIR competition, the strong synergy between neural networks and formal languages is at the core of several papers in these proceedings: from investigating counting behaviors in neural models to tackling transductions and active automata learning with neural networks. Some papers in these proceedings push further our understanding of classical, fundamental and challenging problems in Grammatical Inference. This includes extending distributional learning from positive data and membership queries, the problem of identification of substitutable context-free languages over infinite alphabets, inferring minimalist lexicon from SMT models, novel approaches to learning syntactic monoids and inferring regular expressions.

Honorary chair

- Mohammed Regragui, Dean of the Faculty of Sciences in Rabat.

ICGI Conference Chairs

- François Coste, Inria Rennes, France
- Faissal Ouardi, Faculty of Sciences, Mohammed V University in Rabat, Morocco
- Guillaume Rabusseau, University of Montreal – Mila, Canada

ICGI & TAYSIR Program Committee

We thank the members of the program committee for their careful evaluation of the submissions.

- Borja Balle, Deepmind
- Leonor Becerra, Laboratoire d’Informatique et Systèmes, Aix-Marseille University, France
- Johanna Björklund, Umeå University, Sweden
- Alexander Clark, University of Gothenburg, Sweden
- François Coste, Univ Rennes, Inria, CNRS, IRISA, France
- Rémi Eyraud, Université Jean Monnet, France
- Henning Fernau, Univ Trier, Germany
- Annie Foret, IRISA & University of Rennes 1, France
- Robert Frank, Yale University, USA

PREFACE

- Matthias Gallé, cohere.ai
- Jeffrey Heinz, Stony Brook University, USA
- Falk Howar, TU Clausthal / IPSSE, Germany
- Jean-Christophe Janodet, University of Evry, France
- Adam Jardine, Rutgers University, USA
- Tobias Kappé, Open University of the Netherlands & ILLC, University of Amsterdam, The Netherlands
- Clara Lacroce, McGill, Canada
- Aurélien Lemay, INRIA, France
- Tianyu Li, McGill University, Canada
- Damián López, Universitat Politècnica de València, Spain
- William Merrill, New York University, USA
- Joshua Moerman, Open University of the Netherlands, The Netherlands
- Faissal Ouardi, Mohammed V University in Rabat, Morocco
- Guillaume Rabusseau, Montreal University – Mila, Canada
- Jonathan Rawski, Stony Brook University, USA
- Matteo Sammartino, Royal Holloway University of London, University College London, United Kingdom
- Ute Schmid, University of Bamberg, Germany
- Jose M.Sempere, Universitat Politècnica de València, Spain
- Chihiro Shibata, Hosei University, Japan
- Olgierd Unold, Wroclaw University of Science and Technology, Poland
- Sicco Verwer, Delft University of Technology, The Netherlands
- Gail Weiss, EPFL, Switzerland
- Wojciech Wiczorek, University of Bielsko-Biala, Poland
- Ryo Yoshinaka, Tohoku University, Japan
- Menno van Zaanen, North West University, South of Africa