

# Proceedings of the 13th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2024)



Linköping Electronic Conference Proceedings Nr. 211

eISSN 1650-3740 (Online)  
ISSN 1650-3686 (Print)  
ISBN 978-91-8075-774-4 (Print)

2024



Proceedings of the

# 13th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2024)

edited by

Thomas Gaillat, Cyriel Mallart, Fabienne Moreau, Jen-Yu Li, Griselda Drouet,  
David Alfter, Elena Volodina and Arne Jönsson

Proceedings and all papers therein  
published under a CC BY 4.0 license:  
<https://creativecommons.org/licenses/by/4.0>

Front cover image by ykaiavu  
via Pixabay

Linköping Electronic Conference Proceedings Nr. 211

eISSN 1650-3740 (Online)  
ISSN 1650-3686 (Print)  
ISBN 978-91-8075-774-4 (Print)

2024



## Preface

The workshop series on Natural Language Processing (NLP) for Computer-Assisted Language Learning (NLP4CALL) is a meeting place for researchers working on the integration of Natural Language Processing and Speech Technologies in CALL systems and exploring the theoretical and methodological issues arising in this connection. The latter includes, among others, the integration of insights from Second Language Acquisition (SLA) research, and the promotion of “Computational SLA” through setting up Second Language research infrastructures.

The intersection of Natural Language Processing (or Language Technology / Computational Linguistics) and Speech Technology with Computer-Assisted Language Learning (CALL) brings “understanding” of language to CALL tools, thus making CALL intelligent. This fact has given the name for this area of research –Intelligent CALL, or for short, ICALL. As the definition suggests, apart from having excellent knowledge of Natural Language Processing and/or Speech Technology, ICALL researchers need good insights into second language acquisition theories and practices, as well as knowledge of second language pedagogy and didactics. This workshop therefore invites a wide range of ICALL-relevant research, including studies where NLP-enriched tools are used for testing SLA and pedagogical theories, and vice versa, where SLA theories, pedagogical practices or empirical data are modeled in ICALL tools. The NLP4CALL workshop series is aimed at bringing together competences from these areas for sharing experiences and brainstorming around the future of the field.

### We invited submissions:

- that describe research directly aimed at ICALL
- that demonstrate actual or discuss the potential use of existing Language and Speech Technologies or resources for language learning
- that describe the ongoing development of resources and tools with potential usage in ICALL, either directly in interactive applications, or indirectly in materials, application, or curriculum development, e.g. Large Language Model exploitation, learning material generation, assessment of learner texts and responses, individualized learning solutions, provision of feedback
- that discuss challenges and/or research agenda for ICALL
- that describe empirical studies on language learner data

In this edition of the workshop a special focus was given to systems relying on AI trained for ICALL tasks. This included, but not only, fine tuning Large Language Models (LLMs) and supervised-learning methods based on learning analytics. Issues related to data in SLA and learner corpus such as collection and feature extraction were also welcome. We encouraged paper presentations and software demonstrations describing the above-mentioned themes for all languages.

### Invited speakers

This year, we had the pleasure to welcome two invited speakers: Helen Yannakoudakis (King’s College London) and Kristopher Kyle (University of Oregon).

**Helen Yannakoudakis** is an Assistant Professor at King’s College London and Affiliated Staff at the University of Cambridge. She is also a Turing Fellow and a Fellow of the Higher Education Academy. Helen is working on machine learning for natural language processing

with a focus on few-shot learning, lifelong learning, multilingual NLP, and societal and health applications. Helen’s work has been deployed under the Cambridge brand (Write & Improve), and has been published in leading venues in the field such as NeurIPS and ACL. She has received funding awards from industry and academia, has served as a keynote speaker and a panelist, and has won international competitions such as the NeurIPS 2020 Hateful Memes Challenge. Among others, she has been invited for spotlight interviews (e.g., DrivenData) and comments by media channels such as Reuters and TechCrunch. Recently, she was invited to stay at Windsor Castle to talk about AI in a two-day consultation on threats and opportunities.

In her talk, Helen Yannakoudakis focused on the potential for integrating large language models (LLMs) into AI-powered language teaching and assessment systems. She explored various research areas including content creation, automated grading, and grammatical error correction, while also addressing the risks and ethical concerns surrounding the use of generative AI in language learning technology. Further, she highlighted the need for further research to better understand the strengths and limitations of LLMs and to address foreseeable risks such as misinformation and harmful bias, and explored several directions for future work.

**Kristopher Kyle** is an Associate Professor of Linguistics and Director of the Learner Corpus Research and Applied Data Science Lab. His research interests include natural language processing, corpus linguistics, second language writing, second language assessment, and second language development. (Norwegian Computing Center).

In his talk, Kristopher Kyle provided a brief overview of the use of natural language processing in research related to language learning and assessment over the past 50 years, culminating in the advent of large language models (LLMs). He then briefly discussed some recent (mis)uses of LLMs in language learning and assessment research. He argued that while some black-box LLM-based systems can achieve results with reasonable metrics, such systems tend to have particularly weak validity arguments. He then argued for the development and use of LLM-based systems that increase the construct validity of common CALL applications such as automated evaluation and feedback systems and introduced some working examples of these systems.

## Previous workshops

This workshop follows a series of workshops on NLP4CALL organized by the NEALT Special Interest Group on Intelligent Computer-Assisted Language Learning (SIG-ICALL<sup>1</sup>). The workshop series has previously been financed by the Center for Language Technology at the University of Gothenburg, the SweLL project<sup>2</sup>, the Swedish Research Council’s conference grant, Språkbanken Text<sup>3</sup>, L2 profiling project<sup>4</sup>, itec<sup>5</sup>, the CENTAL<sup>6</sup> and the Analytics for Language Learning (A4LL) project<sup>7</sup> at LIDILE - Univ Rennes.

Submissions to the thirteen workshop editions have targeted a wide range of languages, ranging from well-resourced languages (Chinese, German, English, French, Portuguese, Russian, Spanish) to lesser-resourced languages (Erzya, Arabic, Estonian, Irish, Komi-Zyrian, Meadow Mari, Saami, Udmurt, Võro). Among these, several Nordic languages have been targeted, namely Danish, Estonian, Finnish, Icelandic, Norwegian, Saami, Swedish and Võro. The wide scope of the workshop is also evident in the affiliations of the participating authors as illustrated in Table 1.

---

<sup>1</sup><https://spraakbanken.gu.se/en/research/themes/icall/sig-icall>

<sup>2</sup><https://spraakbanken.gu.se/en/projects/swell>

<sup>3</sup><https://spraakbanken.gu.se>

<sup>4</sup><https://spraakbanken.gu.se/en/projects/l2profiles>

<sup>5</sup><https://itec.kuleuven-kulak.be>

<sup>6</sup><https://cental.uclouvain.be>

<sup>7</sup><https://sites-recherche.univ-rennes2.fr/lidile/articles/a4all/>

<b>Country</b>	<b>Count</b>	<b>Country</b>	<b>Count</b>
Algeria	1	Japan	7
Australia	2	Lithuania	1
Belgium	18	Netherlands	4
Canada	4	Norway	16
Cyprus	3	Portugal	6
Czech Republic	1	Romania	1
Denmark	5	Russia	10
Egypt	1	Slovakia	1
Estonia	3	Spain	5
Finland	15	Sweden	82
France	29	Switzerland	13
Germany	130	UK	23
Iceland	6	Uruguay	5
Ireland	5	US	14
Israel	1	Vietnam	3
Italy	15		

Table 1: NLP4CALL speakers’ and co-authors’ affiliations, 2012–2024

<b>Workshop year</b>	<b>Submitted</b>	<b>Accepted</b>	<b>Acceptance rate</b>
2012	12	8	67%
2013	8	4	50%
2014	13	13	77%
2015	9	6	67%
2016	14	10	72%
2017	13	7	54%
2018	16	11	69%
2019	16	10	63%
2020	7	4	57%
2021	11	6	54%
2022	23	13	56%
2023	18	12	67%
2024	23	19	82%

Table 2: Submissions and acceptance rates, 2012-2024

The acceptance rate has varied between 50% and 82%, the average being 65% (see Table 2). Although the acceptance rate is rather high, the reviewing process has always been very rigorous with two to three double-blind reviews per submission. This indicates that submissions to the workshop have usually been of high quality.

## Program committee

We would like to thank our Program Committee for providing detailed feedback for the reviewed papers:

- Alfter David - University of Gothenburg (Sweden)
- Ar Rouz David - Université Rennes 2 (France)
- Ballier Nicolas - Université Paris Cité (France)
- Balvet Antonio - Université de Lille (France)
- Belan Sophie - Université de Nantes (France)
- Bexte Marie - FernUniversität in Hagen (Germany)
- Bibauw Serge - Université catholique de Louvain (Belgium)
- Caines Andrew - University of Cambridge (United Kingdom)
- Cornillie Frederik - Katholieke Universiteit Leuven (Belgium)
- De Kuthy Kordula - University of Tübingen (Germany)
- Drouet Griselda - Université Rennes 2 (France)
- El Ayari Sarra - CNRS (France)
- Evain Christine - Université Rennes 2 (France)
- Gaillat Thomas - Université Rennes 2 (France)
- Graën Johannes - Universität Zürich (Switzerland)
- Hamilton Clive - Université Paris Cité (France)
- Horbach Andrea - FernUniversität in Hagen (Germany)
- Jönsson Arne - Linköping University (Sweden)
- Laarmann-Quante Ronja - Ruhr-Universität Bochum (Germany)
- Lange Herbert - University of Gothenburg (Sweden)
- Ljunglöf Peter - University of Gothenburg (Sweden)
- Mallart Cyriel - Université Rennes 2 (France)
- Mieskes Margot - Darmstadt University of Applied Sciences / Hochschule Darmstadt (Germany)
- Moreau Fabienne - Université Rennes 2 (France)



- Munoz Garcia Margarita - Université Rennes 2 (France)
- Muñoz Sánchez Ricardo - University of Gothenburg (Sweden)
- Nicolas Lionel - Eurac Research (Italy)
- Pado Ulrike - Hochschule für Technik Stuttgart / University of Applied Sciences of Stuttgart (Germany)
- Paquot Magali - Université catholique de Louvain (Belgium)
- Sarré Cédric - Sorbonne Université (France)
- Stemle Egon - Eurac Research (Italy)
- Vajjala Balakrishna Sowmya - University of Tübingen (Germany)
- Valdez Cristian - Université Paris Cité (France)
- Volodina Elena - University of Gothenburg (Sweden)
- Zesch Torsten - FernUniversität in Hagen (Germany)

We intend to continue this workshop series, which so far has been the only ICALL-related recurring event based in the Nordic countries, Belgium and France. Our intention is to co-locate the workshop series with the two major LT events in Scandinavia, the Swedish Language Technology Conference (SLTC) and the Nordic Conference on Computational Linguistics (NoDaLiDa), thus making this workshop an annual event. Through this workshop, we intend to profile ICALL research in Nordic countries as well as beyond, and we aim at providing a dissemination venue for researchers active in this area.

## Workshop website

[https://nlp4call.github.io/current/past\\_editions.html](https://nlp4call.github.io/current/past_editions.html)

## Workshop organizers

### Université Rennes 2, France

Thomas Gaillat, Cyriel Mallart, Fabienne Moreau, Jen-Yu Li, Griselda Drouet - Linguistique Ingénierie et Didactique des Langues (LIDILE)

### University of Gothenburg, Sweden

David Alfter, Gothenburg Research Infrastructure in Digital Humanities (GRIDH)  
Elena Volodina, Språkbanken Text

### Linköping University, Sweden

Arne Jönsson

## Acknowledgments

We gratefully acknowledge the financial support from the French *Agence Nationale de la Recherche* and the funded project *Analytics for Language Learning A4LL* - ANR-22-CE38-0015.

## Content

Preface	i
<i>Thomas Gaillat, Cyriel Mallart, Fabienne Moreau, Jen-Yu Li, Grisela Drouet, David Alfter, Elena Volodina and Arne Jönsson</i>	
Out-of-the-Box Graded Vocabulary Lists with Generative Language Models: Fact or Fiction?	1
<i>David Alfter</i>	
Investigating Acoustic Correlates of Whisper Scoring for L2 Speech Using Forced alignment with the Italian Component of the ISLE corpus	20
<i>Nicolas Ballier and Adrien Méli</i>	
Leading by Example: The Use of Generative Artificial Intelligence to Create Pedagogically Suitable Example Sentences	33
<i>Jasper Degraeuwe and Patrick Goethals</i>	
Potential of ASR for the study of L2 learner corpora	49
<i>Sarra El Ayari and Zhongjie Li</i>	
Enhancing a multi-faceted verb-centered resource to help a language learner: the case of breton	59
<i>Annie Foret, Erwan Hupel and Pêr Morvan</i>	
Evaluating Automatic Pronunciation Scoring with Crowd-sourced Speech Corpus Annotations	67
<i>Nils Hjortnaes, Daniel Dakota, Sandra Kübler and Francis Tyers</i>	
Opinions Are Buildings: Metaphors in Secondary Education Foreign Language Learning	78
<i>Anna Huelsing and Andrea Horbach</i>	
Investigating strategies for lexical complexity prediction in a multilingual setting using generative language models and supervised approaches	96
<i>Abdelhak Keliou, Mathieu Constant and Christophe Coeur</i>	
Developing a Pedagogically Oriented Interactive Reading Tool with Teachers in the Loops	115
<i>Mihwa Lee, Björn Rudzewitz and Xiaobin Chen</i>	
Developing a Web-Based Intelligent Language Assessment Platform Powered by Natural Language Processing Technologies	126
<i>Sarah Löber, Björn Rudzewitz, Daniela Verratti Souto, Luisa Ribeiro-Flucht and Xiaobin Chen</i>	
Jingle BERT, Jingle BERT, Frozen All the Way: Freezing Layers to Identify CEFR Levels of Second Language Learners Using BERT	137
<i>Ricardo Muñoz Sánchez, David Alfter, Simon Dobník, Maria Szawerna and Elena Volodina</i>	
Generating Contexts for ESP Vocabulary Exercises with LLMs	153
<i>Iglika Nikolova-Stoupak, Serge Bibaww, Amandine Dumont, Françoise Stas, Patrick Watrin and Thomas François</i>	
Automatic Text Simplification: A Comparative Study in Italian for Children with Language Disorders	176
<i>Francesca Padovani, Caterina Marchesi, Eleonora Pasqua, Martina Galletti and Daniele Nardi</i>	

A Conversational Intelligent Tutoring System for Improving English Proficiency of Non-Native Speakers via Debriefing of Online Meeting Transcriptions	187
<i>Juan Antonio Pérez-Ortiz, Miquel Esplà-Gomis, Víctor M. Sánchez-Cartagena, Felipe Sánchez-Martínez, Roman Chernysh, Gabriel Mora-Rodríguez and Lev Berezhnoy</i>	
Evaluating the Generalisation of an Artificial Learner	199
<i>Bernardo Stearns, Nicolas Ballier, Thomas Gaillat, Andrew Simpkin and John P. McCrae</i>	
Semantic Error Prediction: Estimating Word Production Complexity	209
<i>David Strohmaier and Paula Buttery</i>	
GRAMEX: Generating Controlled Grammar Exercises from Various Sources	226
<i>Guillaume Toussaint, Yannick Parmentier and Claire Gardent</i>	
LLM chatbots as a language practice tool: a user study	235
<i>Gladys Tyen, Andrew Caines and Paula Buttery</i>	
Sailing through multiword expression identification with Wiktionary and Linguse: A case study of language learning	248
<i>Till Überrück-Fries, Agata Savary and Agnieszka Dryjańska</i>	