

Proceedings of the 14th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2025)



University of Tartu Library

ISBN 978-9908-53-112-0

2025

Proceedings of the

14th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2025)

edited by

Ricardo Muñoz Sánchez, David Alfter, Elena Volodina and Jelena Kallas

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

Front cover image by Martin Hochreiter (idefixgallier)
via Pixabay

University of Tartu Library

ISBN 978-9908-53-112-0

2025

Preface

The workshop series on Natural Language Processing (NLP) for Computer-Assisted Language Learning (NLP4CALL) is a meeting place for researchers working on integrating 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 inspired the name for this area of research — Intelligent CALL, ICALL for short. 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. Therefore, this workshop invites a wide range of ICALL-relevant research, including studies where NLP-enriched tools are used for testing SLA and pedagogical theories, and those where SLA theories (pedagogical practices or empirical data) are modeled using 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.

Topics of Interest

We invited submissions:

- that describe research directly aimed at ICALL;
- 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. learning material generation, assessment of learner texts and responses, individualized learning solutions, provision of feedback);
- that discuss challenges and/or research agendas for ICALL;
- that describe empirical studies on language learner data; and
- that explore the use of LLMs and Generative AI to develop ICALL tools.

In this edition of the workshop a special focus was given to:

- grammatical error correction and
- the use of pedagogically oriented constructicographic resources (constructicons), with an emphasis on their practical application in ICALL.¹

We encouraged paper presentations and software demonstrations describing the above-mentioned themes primarily, but not exclusively, for the Nordic languages.

A special feature in this year’s workshop was the MultiGEC-2025 shared task on grammatical error correction that was held in connection to the workshop. It featured 12 European languages (Czech, English, Estonian, German, Greek, Icelandic, Italian, Latvian, Russian, Slovene, Swedish

¹By constructicographic resources, we refer to resources that describe various types of constructions associated with specific meanings or functions, ranging from fully schematic and semi-schematic constructions (e.g., those with both fixed and variable elements) to specific lexical expressions.

and Ukrainian) and was organized by the CompSLA working group² as well as over 20 external data providers. A paper describing the shared task and system descriptions from two of the participating teams are included in these proceedings.

Invited speakers

This year, we had the pleasure to welcome two invited speakers: Andrew Caines (University of Cambridge) and Peter Uhrig (Friedrich-Alexander-Universität Erlangen-Nürnberg).

Andrew Caines is a Senior Research Associate based in the Computer Laboratory at the University of Cambridge, U.K. He has been a member of the Institute for Automated Language Teaching & Assessment (ALTA) since its inception in 2013. His research interests relate to education technology for language learning, including corpus creation, automated essay scoring, grammatical error detection and correction, adaptive learning, content creation, and the training of smaller, domain-specific language models. The title of his talk was *The Potential and the Pitfalls of Very Large Language Models for Language Learning Applications*.

Peter Uhrig is professor of Digital Linguistics with a focus on Big Data at Friedrich-Alexander-Universität Erlangen-Nürnberg. His research interests include cognitive linguistics, especially Construction Grammar, collo-phenomena (collocation, collocation), computational and corpus linguistics, and lexicography. He is particularly interested in using large multimodal datasets, data science methods, and machine learning in his work. In addition to his research, Peter Uhrig is committed to creating research infrastructures and open datasets, supporting the broader linguistic community. His work aims to integrate technology with linguistic research, contributing to the evolving field of Digital Linguistics. The title of his talk was *AI-assisted (Pedagogical) Constructicography – Opportunities and Challenges*.

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)³. The workshop series has previously been financed by:

- the Center for Language Technology at the University of Gothenburg;
- the SweLL project;⁴
- the Swedish Research Council’s conference grant, Språkbanken Text;⁵
- the L2 profiling project;⁶
- itec;⁷
- the CENTAL;⁸
- the Analytics for Language Learning (A4LL) project⁹ at LIDILE - Univ Rennes;

²<https://spraakbanken.gu.se/compsla>

³<https://spraakbanken.gu.se/en/research/themes/icall/sig-icall>

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

⁵<https://spraakbanken.gu.se>

⁶<https://spraakbanken.gu.se/en/projects/l2profiles>

⁷<https://itec.kuleuven-kulak.be>

⁸<https://cental.uclouvain.be>

⁹<https://sites-recherche.univ-rennes2.fr/lidile/articles/a4all/>

- the Mormor Karl project;¹⁰ and
- the project "Expanding the scope of a multi-purpose lexicographic resource to grammar and L2 competence".¹¹

Submissions to the fourteen 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.

The acceptance rate has varied between 44% and 82%, the average being 63% (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.

Country	Count	Country	Count
Algeria	1	Japan	7
Australia	2	Lithuania	1
Belgium	20	Netherlands	4
Canada	4	Norway	16
China	5	Poland	1
Cyprus	3	Portugal	8
Czech Republic	1	Romania	1
Denmark	5	Russia	10
Egypt	1	Slovakia	1
Estonia	3	Spain	5
Finland	15	Sweden	87
France	35	Switzerland	15
Germany	135	Ukraine	2
Iceland	6	UK	25
Ireland	5	Uruguay	5
Israel	1	US	15
Italy	15	Vietnam	3

Table 1: NLP4CALL speakers' and co-authors' affiliations, 2012–2025

¹⁰<https://mormor-karl.github.io/>

¹¹<https://eki.ee/prg-1978/>

Workshop year	Submitted	Accepted	Acceptance rate
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%
2025	16	7	44%

Table 2: Submissions and acceptance rates, 2012–2025

Program committee

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

- David Alfter, University of Gothenburg, Sweden
- Serge Bibauw, Universidad Central del Ecuador, Ecuador
- Claudia Borg, University of Malta, Malta
- Christopher Bryant, University of Cambridge, UK
- Andrew Caines, University of Cambridge, UK
- Orphée De Clercq, Ghent University, Belgium
- Kordula de Kuthy, Universität Tübingen, Germany
- Piet Desmet, K.U. Leuven, Belgium
- Thomas François, Université catholique de Louvain, Belgium
- Thomas Gaillat, Université Rennes 2, France
- Andrea Horbach, FernUniversität Hagen, Germany
- Jelena Kallas, Institute of the Estonian Language, Estonia
- Joni Kruijsbergen, Ghent University, Belgium
- Murathan Kurfah, RISE Research Institutes of Sweden, Sweden
- Herbert Lange, University of Gothenburg, Sweden
- Arianna Masciolini, University of Gothenburg, Sweden
- Margot Mieskes, University of Applied Sciences Darmstadt, Germany
- Ricardo Muñoz Sánchez, University of Gothenburg, Sweden
- Lionel Nicolas, EURAC research, Italy
- Ulrike Pado, Hochschule für Technik Stuttgart, Germany
- Magali Paquot, Université catholique de Louvain, Belgium
- Ildikó Pilán, Norwegian Computing Center, Norway
- Gerold Schneider, University of Zurich, Switzerland
- Maria Irena Szawerna, University of Gothenburg, Sweden
- Irina Temnikova, Big Data for Smart Society Institute (GATE)
- Sowmya Vajjala, National Research Council, Canada
- Elena Volodina, University of Gothenburg, Sweden

- Torsten Zesch, FernUniversität Hagen, Germany
- Robert Östling, Stockholm University, Sweden

We intend to continue this workshop series, which so far has been the only ICALL-related recurring event based in the Nordic countries. 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://spraakbanken.gu.se/en/research/themes/icall/nlp4call-workshop-series/nlp4call2025>

Workshop organizers

- Ricardo Muñoz Sánchez, Språkbanken Text, University of Gothenburg, Sweden
- David Alfter, Gothenburg Research Infrastructure in Digital Humanities (GRIDH), University of Gothenburg, Sweden
- Elena Volodina, Språkbanken Text, University of Gothenburg, Sweden
- Jelena Kallas, Institute of the Estonian Language, Estonia

Acknowledgments

The 2025 edition of this workshop was supported jointly by:

- The project *Expanding the scope of a multi-purpose lexicographic resource to grammar and L2 competence*,¹² funded by the Estonian Research Council¹³ grant (PRG 1978).
- The project *Grandma Karl is 27 years old: Automatic pseudonymization of research data*¹⁴ with the Swedish Research Council¹⁵ grant with funding number 2022-02311.
- The research infrastructure Språkbanken,¹⁶ jointly funded by its 10 partner institutions and the Swedish Research Council (2018–2024; dnr 2017-00626)

¹²<https://eki.ee/prg-1978/>

¹³<https://etag.ee/en/>

¹⁴<https://mormor-karl.github.io/>

¹⁵<https://www.vr.se/english.html>

¹⁶<https://spraakbanken.gu.se/>

Content

Preface	i
<i>Ricardo Muñoz Sánchez, David Alfter, Elena Volodina and Jelena Kallas</i>	
The MultiGEC-2025 Shared Task on Multilingual Grammatical Error Correction at NLP4CALL	1
<i>Arianna Masciolini, Andrew Caines, Orphée De Clercq, Joni Kruijsbergen, Murathan Kurfah, Ricardo Muñoz Sánchez, Elena Volodina and Robert Östling</i>	
Lattice @MultiGEC-2025: A Spitful Multilingual Language Error Correction System Using LLaMA	34
<i>Olga Seminck, Yoann Dupont, Mathieu Dehouck, Qi Wang, Noé Durandard and Margo Novikov</i>	
UAM-CSI at MultiGEC-2025: Parameter-efficient LLM Fine-tuning for Multilingual Grammatical Error Correction	42
<i>Ryszard Staruch</i>	
Interpretable Machine Learning for Societal Language Identification: Modeling English and German Influences on Portuguese Heritage Language	50
<i>Soroosh Akef, Detmar Meurers, Amália Mendes and Patrick Rebuschat</i>	
A prototype authoring tool for editing authentic texts using LLMs to increase support for contextualised L2 grammar practice	63
<i>Stephen Bodnar</i>	
PIRLS Category-specific Question Generation for Reading Comprehension	72
<i>Yin Poon, Qiong Wang, John S. Y. Lee, Yu Yan Lam and Samuel Kai Wah Chu</i>	
Investigating Linguistic Abilities of LLMs for Native Language Identification	81
<i>Ahmet Yavuz Ulusu and Gerold Schneider</i>	

