NEALT Proceedings Series Vol. 53

Proceedings of the 12th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2023)

Proceedings of the

12th Workshop on Natural Language Processing for Computer Assisted Langauge Learning (NLP4CALL 2023)

edited by
David Alfter, Elena Volodina, Thomas François, Arne Jönsson and Evelina Rennes

Front cover image by Peggychoucair
via Pixabay

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 and 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. 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 is given to work done on error detection/ correction and feedback generation. 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 a shared task on grammatical error detection that was held in connection to the workshop: the MultiGED shared task on token-level error detection for L2 Czech, English, German, Italian and Swedish, organized by the Computational SLA working group. System descriptions from participating teams are included in these proceedings.

Invited speakers

This year, we had the pleasure to welcome two invited speakers: Marije Michel (University of Groningen) and Pierre Lison (Norwegian Computing Center).

Marije Michel is chair of Language Learning at Groningen University in the Netherlands. Her research and teaching focus on second language acquisition and processing with specific

attention to task-based language pedagogy, digitally-mediated interaction and writing in a second language.

In her talk, *TELL: Tasks Engaging Language Learners*, she reviewed the most important principles of designing engaging learning tasks, highlighted examples of practice-induced L2 research using digital tools, and showcased some of her own work on task design for L2 learning during digitally mediated communication and L2 writing.

Pierre Lison is a senior researcher at the Norwegian Computing Center, a research institute located in Oslo and conducting research in computer science, statistical modelling and machine learning. Pierre's research interests include privacy-enhancing NLP, spoken dialogue systems, multilingual corpora and weak supervision. Pierre currently leads the CLEANUP project on data-driven models for text sanitization. He also holds a part-time position as associate professor at the University of Oslo.

In this talk, *Privacy-enhancing NLP: a primer*, he discussed the privacy concerns associated with personal data in text documents, particularly in the context of Computer-Assisted Language Learning. He highlighted the presence of lexical and grammatical errors that can inadvertently reveal the author's identity and discussed privacy-enhancing techniques to mitigate these risks. These techniques include text sanitization, text rewriting, and privacy-preserving training. He also presented their own research on data-driven text sanitization, which incorporates explicit measures of privacy risks. Furthermore, he introduced the Text Anonymization Benchmark (TAB) as a tool for evaluating such methods.

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[1]). The workshop series has previously been financed by the Center for Language Technology at the University of Gothenburg, the SweLL project[2], the Swedish Research Council's conference grant, Språkbanken Text[3], L2 profiling project[4], itec[5] and the CENTAL[6].

Submissions to the twelve 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 50% and 77%, 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.

---

[1] https://spraakbanken.gu.se/en/research/themes/icall/sig-icall
[2] https://spraakbanken.gu.se/en/projects/swell
[3] https://spraakbanken.gu.se
[4] https://spraakbanken.gu.se/en/projects/l2profiles
[5] https://itec.kuleuven-kulak.be
[6] https://cental.uclouvain.be

Country	Count	Country	Count
Algeria	1	Japan	7
Australia	2	Lithuania	1
Belgium	10	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	4
Finland	15	Sweden	78
France	10	Switzerland	13
Germany	110	UK	18
Iceland	6	Uruguay	5
Ireland	2	US	8
Israel	1	Vietnam	3
Italy	11		

Table 1: NLP4CALL speakers' and co-authors' affiliations, 2012–2023

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%

Table 2: Submissions and acceptance rates, 2012-2023

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

- António Branco, Universidade de Lisboa, Portugal

- Andrew Caines, University of Cambridge, UK

- Xiaobin Chen, Universität Tübingen, Germany

- Frederik Cornillie, University of Leuven, Belgium

- Kordula de Kuthy, Universität Tübingen, Germany

- Piet Desmet, University of Leuven, Belgium

- Thomas François, Université catholique de Louvain, Belgium

- Thomas Gaillat, Université Rennes 2, France

- Johannes Graën, University of Zurich, Switzerland

- Andrea Horbach, FernUniversität Hagen, Germany

- Arne Jönsson, Linköping University, Sweden

- Ronja Laarmann-Quante, FernUniversität Hagen, Germany

- Herbert Lange, University of Hamburg, Germany

- Peter Ljunglöf, University of Gothenburg, Sweden and Chalmers Institute of Technology, Sweden

- Margot Mieskes, University of Applied Sciences Darmstadt, Germany

- Lionel Nicolas, EURAC research, Italy

- Ulrike Pado, Hochschule für Technik Stuttgart, Germany

- Magali Paquot, Université catholique de Louvain, Belgium

- Evelina Rennes, Linköping University, Sweden

- Egon Stemle, EURAC research, Italy

- Francis M. Tyers, Indiana University Bloomington, US

- Sowmya Vajjala, National Research Council, Canada

- Elena Volodina, University of Gothenburg, Sweden

- Zarah Weiss, Universität Tübingen, Germany

Proceedings of the 12th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2023)

iv

- Torsten Zesch, FernUniversität Hagen, Germany

- Ramon Ziai, Universität Tübingen, 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 weell 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/nlp4call2023

Workshop organizers

- David Alfter, Gothenburg Research Infrastructure in Digital Humanities (GRIDH), University of Gothenburg, Sweden

- Elena Volodina, Språkbanken Text, University of Gothenburg, Sweden

- Thomas François, Cental, Université catholique de Louvain, Belgium

- Arne Jönsson, Department of Computer and Information Science, Linköping University, Sweden

- Evelina Rennes, Department of Computer and Information Science, Linköping University, Sweden

Acknowledgments

---

[7]https://spraakbanken.gu.se/en/projects/mormor-karl
[8]https://spraakbanken.gu.se/
[9]https://www.huminfra.se/

Content

Proceedings of the 12th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2023)

vi