Proceedings of the 11th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2022)

Proceedings of the

11th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2022)

edited by
David Alfter, Elena Volodina, Thomas François, Piet Desmet,
Frederik Cornillie, Arne Jönsson and Evelina Rennes

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 modelled 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

This year a special focus was given to work done on second language vocabulary and grammar profiling, as well as the use of crowdsourcing for creating, collecting, and curating data in NLP projects. 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 is the *research notes* session. This session included short talks about PhD projects and ongoing unfinished research that collaborating teams were eager to discuss with the community and get feedback. We tested this feature for the second time with an intention to evaluate its impact and utility for future uses. This time around, we circulated a separate call for expression of interest.

This year, we had the pleasure to welcome two invited speakers: Christopher Bryant (Reverso and University of Cambridge) and Marije Michel (University of Groningen).

Dr **Christopher J. Bryant** is a Principal Applied Research Scientist at the AI-powered translation and language tool service Reverso, and a Research Associate in the Institute for Automated Language Teaching and Assessment (ALTA) at the University of Cambridge. His main research interests include automatic grammatical error detection and correction (GED/GEC), automatic corpus annotation, and the robust evaluation of grammatical error correction systems, along with computer-aided language learning in general. He completed his PhD on "Automatic annotation for grammatical error correction" in 2019 and is the lead developer of the associated ERRor ANnotation Toolkit (ERRANT) which is

widely used to benchmark progress in the field. He led the most recent shared task on GEC in 2019 (BEA-2019) and currently researches artificial error generation for GEC and develops the Ginger grammar checker at Reverso.

In his talk, *The Evolution of Automatic Grammatical Error Correction*, he provided an overview of the field and introduced the datasets, approaches and evaluation methods that are commonly used to build Grammatic Error Correction systems. He concluded with recent trends and remaining challenges.

Dr **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.

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] and the L2 profiling project[4]. This year's workshop is jointly financed by itec[5] and the CENTAL[6].

Submissions to the eleven 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.

Country	2012-2022 (# speaker/co-author affiliations)
Algeria	1
Australia	2
Belgium	9
Canada	4
Cyprus	2
Denmark	3
Egypt	1
Estonia	3
Finland	10
France	10
Germany	103
Iceland	6

Proceedings of the 11th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2022)

ii

Ireland	2
Italy	7
Japan	7
Lithuania	1
Netherlands	4
Norway	13
Portugal	6
Romania	1
Russia	10
Slovakia	1
Spain	4
Sweden	71
Switzerland	11
UK	16
US	8

Table 1. NLP4CALL speakers' and co-authors' affiliations, 2012-2022

The acceptance rate has varied between 50% and 77%, 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.

Workshop year	Submitted	Accepted	Acceptance rate
2012	12	8	67%
2013	8	4	50%
2014	13	10	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%

Table 2: Submissions and acceptance rates, 2012-2021

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

- David Alfter, Université catholique de Louvain, Belgium and 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
- Johannes Graën, University of Zurich, Switzerland
- Andrea Horbach, FernUniversität Hagen, Germany
- Arne Jönsson, Linköping University, Sweden

Proceedings of the 11th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2022)

iii

- 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
- Rodrigo Wilkens, Université catholique de Louvain, Belgium
- 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-relevant recurring event based in the Nordic countries. Our intention is to co-locate the workshop series with the two major LT events in Scandinavia, SLTC (the Swedish Language Technology Conference) and NoDaLiDa (Nordic Conference on Computational Linguistics), 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/forskning/teman/icall/nlp4call-workshop-series/nlp4call2022

Workshop organizers

David Alfter[1,2], Elena Volodina[2], Thomas François[1], Piet Desmet[3], Frederik Cornillie[3], Arne Jönsson[4], Evelina Rennes[4]

[1] Cental, Université catholique de Louvain, Belgium

[2] Språkbanken, University of Gothenburg, Sweden

[3] Itec, Department of Linguistics at KU Leuven & imec, Belgium

[4] Department of Computer and Information Science, Linköping University, Sweden

Acknowledgements

Proceedings of the 11th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2022)

iv

Content

Proceedings of the 11th Workshop on Natural Language Processing for Computer Assisted Language Learning (NLP4CALL 2022)

v