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

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

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

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-ICALL¹). The workshop series has previously been financed by the Center for Language Technology at the University of Gothenburg, the SweLL project², the Swedish Research Council’s conference grant, Språkbanken Text³, L2 profiling project⁴, itec⁵ and the CENTAL⁶.

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.

---

¹https://spraakbanken.gu.se/en/research/themes/icall/sig-icall
²https://spraakbanken.gu.se/en/projects/swell
³https://spraakbanken.gu.se
⁴https://spraakbanken.gu.se/en/projects/l2profiles
⁵https://itec.kuleuven-kulak.be
⁶https://cental.uclouvain.be
<table>
<thead>
<tr>
<th>Country</th>
<th>Count</th>
<th>Country</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>1</td>
<td>Japan</td>
<td>7</td>
</tr>
<tr>
<td>Australia</td>
<td>2</td>
<td>Lithuania</td>
<td>1</td>
</tr>
<tr>
<td>Belgium</td>
<td>10</td>
<td>Netherlands</td>
<td>4</td>
</tr>
<tr>
<td>Canada</td>
<td>4</td>
<td>Norway</td>
<td>16</td>
</tr>
<tr>
<td>Cyprus</td>
<td>3</td>
<td>Portugal</td>
<td>6</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>1</td>
<td>Romania</td>
<td>1</td>
</tr>
<tr>
<td>Denmark</td>
<td>5</td>
<td>Russia</td>
<td>10</td>
</tr>
<tr>
<td>Egypt</td>
<td>1</td>
<td>Slovakia</td>
<td>1</td>
</tr>
<tr>
<td>Estonia</td>
<td>3</td>
<td>Spain</td>
<td>4</td>
</tr>
<tr>
<td>Finland</td>
<td>15</td>
<td>Sweden</td>
<td>78</td>
</tr>
<tr>
<td>France</td>
<td>10</td>
<td>Switzerland</td>
<td>13</td>
</tr>
<tr>
<td>Germany</td>
<td>110</td>
<td>UK</td>
<td>18</td>
</tr>
<tr>
<td>Iceland</td>
<td>6</td>
<td>Uruguay</td>
<td>5</td>
</tr>
<tr>
<td>Ireland</td>
<td>2</td>
<td>US</td>
<td>8</td>
</tr>
<tr>
<td>Israel</td>
<td>1</td>
<td>Vietnam</td>
<td>3</td>
</tr>
<tr>
<td>Italy</td>
<td>11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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

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 Vajjula, National Research Council, Canada
• Elena Volodina, University of Gothenburg, Sweden
• Zarah Weiss, Universität Tübingen, Germany
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/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
We gratefully acknowledge the financial support from the VR funded project Grandma Karl is 27 years old: automatic pseudonymization of research data7, project id 2022-02311_VR.
Elena Volodina has been supported by the Swedish Language Bank Språkbanken Text8 and by HumInfra9 through funding from the Swedish Research Council (contracts 2017-00626 and 2021-00176).

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

Preface i
David Alfter, Elena Volodina, Thomas François, Arne Jönsson and Evelina Rennes

MultiGED-2023 shared task at NLP4CALL: Multilingual Grammatical Error Detection 1
Elena Volodina, Christopher Bryant, Andrew Caines, Orphée De Clercq, Jennifer-Carmen Frey, Elizaveta Ershova, Alexandr Rosen and Olga Vinogradova

NTNU-TRH system at the MultiGED-2023 Shared on Multilingual Grammatical Error Detection 17
Lars Bungum, Björn Gambäck and Arild Brandrud Ness

EliCoDe at MultiGED2023: fine-tuning XLM-RoBERTa for multilingual grammatical error detection 24
Davide Colla, Matteo Delsanto and Elisa Di Nuovo

A distantly supervised Grammatical Error Detection/Correction system for Swedish 35
Murathan Kurfalı and Robert Östling

Two Neural Models for Multilingual Grammatical Error Detection 40
Phuong Le-Hong, The Quyen Ngo and Thi Minh Huyen Nguyen

Experiments on Automatic Error Detection and Correction for Uruguayan Learners of English 45
Romina Brown, Santiago Paez, Gonzalo Herrera, Luis Chiruzzo and Aiala Rosá

Sequence Tagging in EFL Email Texts as Feedback for Language Learners 53
Yuning Ding, Ruth Trüb, Johanna Fleckenstein, Stefan Keller and Andrea Horbach

Speech Technology to Support Phonics Learning for Kindergarten Children at Risk of Dyslexia 63
Stine Fuglsang Engmose and Peter Juel Henrichsen

On the relevance and learner dependence of co-text complexity for exercise difficulty 71
Tanja Heck and Detmar Meurers

Manual and Automatic Identification of Similar Arguments in EFL Learner Essays 85
Ahmed Mousa, Ronja Laarmann-Quante and Andrea Horbach

DaLAJ-GED - a dataset for Grammatical Error Detection tasks on Swedish 94
Elena Volodina, YousufAliMohammed, Aleksandrs Berdicevskis, Gerlof Bouma and Joey Öhman

Automated Assessment of Task Completion in Spontaneous Speech for Finnish and Finland Swedish Language Learners 102
Ekaterina Voskoboinik, Yaroslav Getman, Ragheb Al-Ghezi, Mikko Kurimo and Tamas Grosz