

# Crowdsourcing for language learning and linguistic resource creation

*Lionel NICOLAS*

Eurac Research, Institute for Applied Linguistics, Bolzano

*Verena LYDING*

Eurac Research, Institute for Applied Linguistics, Bolzano

The current special issue of the journal *Slovenščina 2.0* focuses on the newly explored combination of crowdsourcing, language learning and linguistic resource creation. It contains five articles and one project report, providing insightful discussions on several aspects of this combination, as well as results which help us understand its versatile potential and the challenges to address in order to better exploit it.

This issue is directly related to the European Network for Combining Language Learning with Crowdsourcing Techniques (enetCollect) and constitutes the first milestone of its follow-up initiative the D4Collect Dariah Working Group. EnetCollect was a large network project funded as a COST Action which started in February 2017 with the objective of creating a research and innovation community to explore the subject. After the end of enetCollect's funding period, a Dariah Working Group called D4Collect was created to keep the community together and continue to foster and coordinate further research on language

---

*Nicolas, L., Lyding, V.: Crowdsourcing for language learning and linguistic resource creation. Slovenščina 2.0, 10(2): 1–4.*

1.20 Uvodnik / Editorial

DOI: <https://doi.org/10.4312/slo2.0.2022.2.1-4>

<https://creativecommons.org/licenses/by-sa/4.0/>



learning combined with crowdsourcing techniques. This special issue features an international set of 14 authors from 13 different countries (see the author list below), and most of the discussions and results presented are the outcome of efforts and collaborations either initiated or intensified in the context of enetCollect. This issue of the journal was also graciously supported in the review process by another international set of 10 enetCollect members from 10 different countries (see the reviewer list below). As guest editors of this special issue, we would like to thank the authors and reviewers for their contribution and, at the same time, express our gratitude to the main editors of *Slovenščina 2.0* for offering to center the 2022 special edition on the combination of crowdsourcing, language learning and linguistic resource creation.

The first article by Volodina, Alfter and Lindström Tiedemann on the Swedish language explores the concepts of core vocabulary learning, non-expert crowdsourcing, CEFR assignments and comparative judgments. More precisely, it investigates the theoretical and practical issues connected to identifying core vocabulary at different levels of linguistic proficiency using statistical approaches combined with crowdsourcing. At the same time, the authors investigate whether crowdsourcing second language learners' rankings can be used in a comparative judgment setting for assigning CEFR levels to unseen vocabulary.

In the second article by Zingaro Kuhn, Arhar Holdt, Kosem Tiberius, Koppel and Zviel-Girshin, the authors describe the development and use of a game-with-a-purpose (GWAP) to crowdsource a pedagogical corpus of example sentences showcasing different types of problems (sensitive content, offensive language, structural problems) for Dutch, Estonian, Slovene and Brazilian Portuguese. They provide a design based on initial experiments focusing on the crowdsourcing suitability of a GWAP in which players identify and classify problematic sentences, and point out problematic excerpts. They also present the methodology for data preparation in terms of source corpora selection, pedagogically oriented GDEX (Good Dictionary EXamples) configurations, and the creation of lemma lists, with a special focus on common and language-dependent decisions.

The third article by Graën explores the concept of generating language learning exercises from parallel corpora and crowdsourcing the actions of users (both learners and teachers) to improve the quality of the corpora. The article provides a blueprint for such a generation mechanism and details three main challenges to tackle when implementing it. It also discusses the fact that, through triangulation, user actions can be transferred to language pairs other than the original ones if multiparallel corpora are used as a source to generate exercises.

The fourth article by Hatipoğlu, Delibegović Džanić, Gajek and Miloshevska addresses the awareness and popularity of crowdsourcing solutions among language learners before and after the COVID-19 pandemic. More precisely, they show that the changes brought about by COVID-19 to educational systems worldwide noticeably impacted language learners' habits and attitudes towards the use of crowdsourcing materials in Turkey, Bosnia and Herzegovina, the Republic of North Macedonia and Poland. They also discuss how, among other factors, students' reduced interactions with teachers and peers added to their workload, and that the lack of support on the part of institutions led them to take more responsibility for their own learning.

The fifth article, by Gajek, shows how the eTwinning European framework, originally designed to facilitate collaboration among schools in Europe and beyond, can be considered as an extensive educational crowdsourcing activity, and demonstrates that teachers can effectively use crowdsourcing in educational practice. Accordingly, the author undertakes two kinds of analyses: a global analysis of the features of the framework in light of crowdsourcing principles, and a local analysis of a selection of outstanding projects submitted for evaluation for national awards in Poland.

In the final article, Nicolas and Lyding review the enetCollect project itself and examine to what extent it met its network-oriented and research-oriented goals that aimed at nurturing a new research and innovation community in order to foster the long-term exploration of the combination of crowdsourcing and language learning. They also introduce the D4Collect Dariah Group that they created to follow-up on enetCollect.

### **Author list**

- David ALFTER, Université Catholique de Louvain, Belgium
- Špela ARHAR HOLDT, University of Ljubljana, Slovenia
- Nihada DELIBEGOVIĆ DŽANIĆ, University of Tuzla, Bosnia and Herzegovina
- Elżbieta GAJEK, University of Warsaw, Poland
- Johannes GRAËN, University of Zurich, Switzerland
- Çiler HATIPOĞLU, Middle East Technical University, Turkey
- Kristina KOPPEL, Institute of the Estonian Language, Estonia
- Iztok KOSEM, University of Ljubljana, Slovenia
- Therese LINDSTRÖM TIEDEMANN, University of Helsinki, Finland
- Lina MILOSHEVSKA, University of Information Science and Technology, North Macedonia
- Carole TIBERIUS, Institute for Dutch Lexicology, Netherlands
- Elena VOLODINA, University of Gothenburg, Sweden
- Tanara ZINGANO KUHN, University of Coimbra, Portugal
- Rina ZVIEL-GIRSHIN, Ruppin Academic Center, Israël

### **Reviewer List**

- Petra BAGO, University of Zagreb, Croatia
- Andrew CAINES, University of Cambridge, United Kingdom
- Gülşen ERYIĞIT, Istanbul Technical University, Turkey
- Thomas FRANÇOIS, Catholic University of Louvain, Belgium
- Andrea HORBACH, University of Hagen, Germany
- Sviatlana KARPAVA, University of Cyprus, Cyprus
- Ivana KONJIK, Institute for the Serbian Language of SASA, Serbia
- Kristina KOPPEL, Institute of the Estonian Language, Estonia
- Julia OSTANINA-OLSZEWSKA, Pedagogical University of Kraków, Poland
- Edina ŠPAGO-ĆUMURIJA, Dzemal Bijedic University of Mostar, Bosnia and Herzegovina