

LREC-COLING 2024

**The Second International Workshop Towards Digital
Language Equality (TDLE): Focusing on Sustainability
(LREC-COLING 2024)**

Workshop Proceedings

Editors

Federico Gaspari, Joss Moorkens, Itziar Aldabe, Aritz
Farwell, Begoña Altuna, Stelios Piperidis, Georg Rehm and
German Rigau

25 May, 2024
Torino, Italia

Proceedings of the Second International Workshop Towards Digital Language Equality (TDLE): Focusing on Sustainability

Copyright ELRA Language Resources Association (ELRA), 2024
These proceedings are licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-48-7
ISSN 2951-2093 (COLING); 2522-2686 (LREC)

Jointly organized by the ELRA Language Resources Association
and the International Committee on Computational Linguistics

Message from the Program Chairs

This volume includes the papers that were presented at the *Second International Workshop Towards Digital Language Equality (TDLE): Focusing on Sustainability*, co-located with the *2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* in Turin, Italy, on 25 May 2024. The event was a follow-up to the first *Workshop Towards Digital Language Equality (TDLE)*, held at the previous LREC Conference in Marseille (France) in June 2022.

We were very pleased with the continued interest in this workshop series from a broad and diverse community of developers, creators, vendors, distributors, brokers, users, evaluators and researchers of Language Resources and Technologies (LRTs) in various (combinations of) languages. Encouraged by this support as well as by the increasing body of work, research, publications, projects and initiatives, we are continuing our efforts dating back several years to promote Digital Language Equality, focusing this time on the crucial issue of sustainability. We believe that this is of great importance and interest to scholars and researchers, industry and commercial players, decision- and policy-makers at various local, regional, national and supranational levels, as well as to the broader public, at a time of unprecedented and fast-paced technological progress, especially in terms of language-centric applications and tools powered by artificial intelligence (AI).

Following a thorough peer-review process, six papers were eventually accepted for presentation at the workshop. The sequence of the papers gathered in these proceedings follows the rationale of going from the broad to the particular, with the first three contributions discussing various aspects and challenges of the ambitious path towards Digital Language Equality, and the remaining three papers focusing on a range of issues having to do with sustainability.

We are very grateful to the authors who submitted their paper proposals for review to the workshop – unfortunately, due to time constraints, it was not possible to accept all of them for presentation and inclusion in the proceedings. Many thanks are due to the Organizing Committee for generously contributing to the successful preparation and running of the workshop, and to the Programme Committee members for sharing their valuable expertise by providing comprehensive and helpful reviews of the submissions that were received. Finally, our heartfelt thanks to the evaluation committee of the general International LREC-COLING 2024 Conference for accepting this workshop as a co-located event, thus recognizing the importance and value to the community of striving for Digital Language Equality and of promoting sustainability with regard to LRTs.

Organizing Committee

Itziar Aldabe (HiTZ Center, UPV/EHU, Spain)
Begoña Altuna (HiTZ Center, UPV/EHU, Spain)
Aritz Farwell (HiTZ Center, UPV/EHU, Spain)
Federico Gaspari (Uni. Naples "Federico II", Italy & ADAPT Centre, DCU, Ireland – co-chair)
Joss Moorkens (SALIS/ADAPT Centre, DCU, Ireland – co-chair)
Stelios Piperidis (ILSP, Athena RC, Greece)
Georg Rehm (DFKI, Germany)
German Rigau (HiTZ Center, UPV/EHU, Spain)

Program Committee

Antonios Anastasopoulos (GMU, USA)
Steven Bird (CDU, Australia)
Fred Blain (Uni. Tilburg, Netherlands)
Franco Cutugno (Uni. Naples "Federico II", Italy)
Bessie Dendrinou (NKUA, Greece & ECSPM, Denmark)
Félix do Carmo (Uni. Surrey, UK)
Annika Grützner-Zahn (DFKI, Germany)
Ana Guerberof-Arenas (Uni. Groningen, Netherlands)
Davyth Hicks (ELEN, Belgium)
Monja Jannet (SALIS/ADAPT, DCU, Ireland)
John Judge (ADAPT, DCU, Ireland)
Dorothy Kenny (SALIS/CTTS/ADAPT, DCU, Ireland)
Sabine Kirchmeier (EFNIL, Luxembourg)
Teresa Lynn (MBZUAI, United Arab Emirates)
Maite Melero (BSC, Spain)
Helena Moniz (Uni. Lisbon, Portugal & EAMT)
Johanna Monti (UniOR, Italy)
Rachele Raus (UniBO, Italy)
Wessel Reijers (Uni. Paderborn, Germany)
Celia Rico Pérez (UCM, Spain)
Dimitar Shterionov (TU, Netherlands)
Carlos S. C. Teixeira (IOTA Localisation Services & Uni. Rovira i Virgili, Spain)
Antonio Toral (Uni. Groningen, Netherlands)
Vincent Vandeghinste (Instituut voor de Nederlandse Taal, Netherlands & KU Leuven, Belgium)

Table of Contents

<i>Surveying the Technology Support of Languages</i> Annika Grützner-Zahn, Federico Gaspari, Maria Giagkou, Stefanie Hegele, Andy Way and Georg Rehm	1
<i>Which Domains, Tasks and Languages are in the Focus of NLP Research on the Languages of Europe?</i> Diego Alves, Marko Tadić and Georg Rehm	18
<i>Fine-Tuning Open Access LLMs for High-Precision NLU in Goal-Driven Dialog Systems</i> Lluís Padró and Roser Saurí	33
<i>Could We Have Had Better Multilingual LLMs if English Was Not the Central Language?</i> Ryandito Diandaru, Lucky Susanto, Zilu Tang, Ayu Purwarianti and Derry Tanti Wijaya .	43
<i>A Language Model Trained on Uruguayan Spanish News Text</i> Juan Pablo Filevich, Gonzalo Marco, Santiago Castro, Luis Chiruzzo and Aiala Rosá ..	53
<i>Environmental Impact Measurement in the MentalRiskES Evaluation Campaign</i> Alba M. Mármol Romero, Adrián Moreno-Muñoz, Flor Miriam Plaza-del-Arco, M. Dolores Molina González and Arturo Montejo-Ráez	61

Workshop Program

Saturday, May 25, 2024

- 14:00–14:30** ***Introduction to the TDLE Workshop***
- 14:30–15:00 *Surveying the Technology Support of Languages*
Annika Grützner-Zahn, Federico Gaspari, Maria Giagkou, Stefanie Hegele,
Andy Way and Georg Rehm
- 15:00–15:30 *Which Domains, Tasks and Languages are in the Focus of NLP Research
on the Languages of Europe?*
Diego Alves, Marko Tadić and Georg Rehm
- 15:30–16:00 *Fine-Tuning Open Access LLMs for High-Precision NLU in Goal-Driven Dia-
log Systems*
Lluís Padró and Roser Saurí
- 16:30–17:00 *Could We Have Had Better Multilingual LLMs if English Was Not the Central
Language?*
Ryandito Diandaru, Lucky Susanto, Zilu Tang, Ayu Purwarianti and Derry
Tanti Wijaya
- 17:00–17:30 *A Language Model Trained on Uruguayan Spanish News Text*
Juan Pablo Filevich, Gonzalo Marco, Santiago Castro, Luis Chiruzzo and
Aiala Rosá
- 17:30–18:00 *Environmental Impact Measurement in the MentalRiskES Evaluation Cam-
paign*
Alba M. Mármol Romero, Adrián Moreno-Muñoz, Flor Miriam Plaza-del-
Arco, M. Dolores Molina González and Arturo Montejo-Ráez