# Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia (EURALI)

# @ LREC 2022

**Date:** Monday, June 20, 2022

**Venue:** Palais du Pharo, Marseille (France)

**Main website:** https://sites.google.com/view/eurali/

**LREC website:** https://lrec2022.lrec-conf.org/en/

---

## Workshop overview and objectives

This workshop will focus on the development of language technology resources and tools for the indigenous, endangered and lesser-resource languages in the Eurasia continent.

In a media-centric world where language technology allows people to break cultural and language barriers, it is important that speakers of endangered and indigenous languages can be empowered to use these technologies to share their knowledge and culture with the world. With the aim of bridging this gap, the goal of this workshop is to increase visibility and promote research for lesser-resourced and underrepresented language communities in Europe and Asia. Through collaboration between NLP researchers, language experts and linguists working for endangered languages in these communities, we aim to create language technology resources that will help to preserve and revive these languages for future generations. Furthermore, the workshop aims to promote the emergence of new methods that benefit linguists, for instance for automation of analysis and validation processes, field linguists, for instance the facilitation of data collection and analysis processes, and computational linguists by developing new techniques necessary for linguistic analysis, development of supervised or weakly-supervised methods for the analysis of poorly written or undocumented languages.

The main objective of the workshop is to create basic resources and develop tools for Eurasiatic languages, including but not limited to the following topics:

  • identifying languages and variants spoken in these regions

   • creating language resources and applications, e.g., sentiment analysis, named entity recognition, and syntactic parsing

  • standardization for endangered languages

  • automatic identification and classification of lexical variation and language varieties

   • adaptation of fundamental NLP tools for these languages, e.g., morphological analysis, taggers and parsers

  • reusability of language resources in NLP applications, e.g., machine translation, POS tagging.

  • machine translation between closely related languages

  • evaluation of language resources and tools when applied to lesser-resourced languages in the same

language families

- corpora, resources, and tools for close related languages

- linguistic and textual similarities among languages in Eurasia

- digitization of endangered languages

- challenges in the creation of language resources and tools from linguistics perspectives

- Linguistics for poorly spoken or undocumented languages

## Submissions

We are seeking submissions under the following category:

Full papers : 8 pages+unlimited reference

Short papers (work in progress) : 4 pages+unlimited reference

Posters (innovative ideas/proposals, a research idea of students) : 4 pages+unlimited reference

Demo (of working online/standalone systems) : 2 pages

Papers must describe original, completed or in progress, and unpublished work. Each submission will be reviewed by three program committee members. The accepted papers will be given up to for full/short paper and poster in the workshop proceedings, and will be presented as an oral presentation or poster.

Papers should be formatted according to the LREC style-sheet, which is provided on the LREC 2022 website (https://lrec2022.lrec-conf.org/en/submission2022/authors-kit/). Please submit papers in PDF format at the START account (the link will be available very soon on the workshop website).

For further information on this initiative, please refer to https://sites.google.com/view/eurali/.

## Important Dates

April 08, 2022: Paper submissions due

May 03, 2022: Paper notification of acceptance

May 23, 2022: Camera-ready papers due

June 20, 2022: Workshop

## Workshop Chair:

Atul Kr. Ojha, National University of Ireland Galway, Ireland

Sina Ahmadi, National University of Ireland Galway, Ireland

Chao-Hong Liu, Potamu Research Ltd, Ireland

John P. McCrae, National University of Ireland Galway, Ireland

**Programme Committee (to be updated):**

Agata Savary, University of Paris-Saclay, France

Alina Karakanta, Fondazione Bruno Kessler (FBK) / University of Trento

Atul Kr. Ojha, National University of Ireland Galway, Irelnad

Bharathi Raja Chakravarthi, National University of Ireland Galway, Irelnad

Bogdan Babych, Heidelberg University, Germany

Chao-Hong Liu, Potamu Research Ltd

Daan vaan Escah, Google

Esha Banerjee, Google, USA

Ekaterina Vylomova, University of Melbourne, Australia

Graham Neubig*, CMU, USA

George Rehm, DFKI GmbH, Germany

John Ortega*, New York University, USA

Jonathan Washington, Swarthmore College, USA

John P. McCrae, National University of Ireland Galway, Irelnad

Joseph Mariani, LIMSI-CNRS, France

Katharina Kann*, University of Colorado at Boulder, USA

Khalid Choukri, ELDA/ELRA, France

Nicoletta Calzolari, CNR-ILC, Italy

Rico Sennrich, University of Zurich, Switzerland

Sina Ahamdi, National University of Ireland Galway, Irelnad

Sunipa Dev, University of California, Los Angeles

Daniel Zeman*, Charles University, Prague

**Identify, Describe, and Share your LRs!**

- Describing your LRs in the LRE Map is now a normal practice in the submission procedure of LREC (introduced in 2010 and adopted by other conferences). To continue the efforts initiated at LREC 2014 about "Sharing LRs" (data, tools, web-services, etc.), authors will have the possibility, when submitting a paper, to upload LRs in a special LREC repository. This effort of

sharing LRs, linked to the LRE Map for their description, may become a new "regular" feature for conferences in our field, thus contributing to creating a common repository where everyone can deposit and share data.

- As scientific work requires accurate citations of referenced work so as to allow the community to understand the whole context and also replicate the experiments conducted by other researchers, LREC 2022 endorses the need to uniquely Identify LRs through the use of the International Standard Language Resource Number (ISLRN, www.islrn.org), a Persistent Unique Identifier to be assigned to each Language Resource. The assignment of ISLRNs to LRs cited in LREC papers will be offered at submission time.