



SIGUL 2022 Workshop
to be held as part of the 13th edition of the
Language Resources and Evaluation Conference (LREC 2022)

1st Call for Papers

Date: 24-25 June 2022

Venue: Palais du Pharo, Marseille (France)

Web Site: <http://sigul-2022.ilc.cnr.it>

Submission Deadline: 11 April 2022

Submission Page: <https://www.softconf.com/lrec2022/SIGUL>

Workshop Description and Objectives

The 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages (SIGUL 2022) will provide a forum for the presentation of cutting edge research in NLP for under-resourced languages to both academic and industry researchers. SIGUL 2022 will carry on the tradition of the CCURL-SLTU (Collaboration and Computing for Under-Resourced Languages – Spoken Language Technologies for Under-resourced languages) Workshop Series, which have been organized since 2008 and, as LREC Workshops, since 2014. As usual, this Workshop spans the research interest areas of less-resourced, under-resourced, endangered, minority and minoritized languages. Since from this year LREC includes a track dedicated specifically to endangered and less resourced languages, the workshop aims to be a venue for networking and discussion as much as for scientific debate.

Over the last years, research in NLP for less-resourced languages has taken momentum. Initiatives and events have flourished, such as AfricaNLP, ComputEL, LT4All, as well as hackatons, toolkits, special interest groups (SIGEL) and journals' special issues. The topic of less-resourced languages has ceased to be niche and has gained space in major conferences such as LREC and ACL (see for instance the ACL 2022 Theme Track on "Language Diversity: from Low-Resource to Endangered Languages").

The multiplication of research interest makes it even more necessary for the community that revolves around less-resourced languages to find opportunities for aggregation and discussion. It is also very important that these occasions leave space for communities and representatives of under-resourced and endangered languages, in order to ensure that the research and development of technological solutions are in line with the needs and demands of those communities, with a view to an open and inclusive research with strong social impact.

Following the long-standing series of previous meetings, the SIGUL venue will provide a forum for the presentation of cutting edge research in NLP for under-resourced languages to both academic and industry researchers and also to offer a venue where researchers in different disciplines and from varied backgrounds can fruitfully explore new areas of intellectual and practical development while honoring their common interest of sustaining less-resourced languages.

Topics include but are not limited to:

- general research on under-resourced languages.
- transfer-learning techniques for under-resourced languages (zero-shot, few-shot training);
- unsupervised and semi-supervised methods to build applications for under-resourced languages;
- use of multilingual pre-trained language models to under-resourced languages;
- speech technologies for under-resourced languages.

We also invite position papers on methodological, ethical or institutional issues.

Joint Session with MWE 2022

SIGUL 2022 will hold a joint session with the Workshop on MultiWord Expressions (MWE 2022). The goal is to foster future synergies that could address scientific challenges in the creation of resources, models and applications to deal with multiword expressions and related phenomena in low-resource scenarios, in accordance with one of the special topics in MWE 2022.

Submission and Publication

Papers need to address less-resourced languages. They can contain an analysis and insight into existing methods and problems; a description of resources; an overview of the literature or of current initiatives, or a combination of the above. Authors must declare if part of the paper contains material previously published elsewhere.

We accept the following typologies of papers:

- research papers;
- position papers, for reflective considerations of methodological, best practice, institutional, etc. issues;
- research posters, for work-in-progress, projects in early stage of development or description of new resources.

All typologies are allowed a maximum of 8 pages, references excluded. The program committee reserves the right to decide whether a paper submitted as a research paper is better suited for a poster presentation.

Accepted papers will be published in online proceedings.

Papers must strictly comply with the LREC stylesheet (<https://lrec2022.lrec-conf.org/en/submission2022/authors-kit>) and be submitted in PDF unprotected format to the following submission page: <https://www.softconf.com/lrec2022/SIGUL>.

Each submission will be reviewed by three programme committee members. In compliance with the LREC rules, papers must **not** be anonymized.

Important Dates

- Paper submission deadline: 11 April 2022
- Notification of acceptance: 3 May 2022
- Camera-ready paper: 23 May 2022
- Workshop date: 24-25 June 2022

Invited Speakers

TBA

Organising Committee

- Maite Melero (Barcelona Supercomputing Center, Spain)
- Sakriani Sakti (NAIST, Japan)
- Claudia Soria (CNR-ILC, Italy)

To contact the organizers, please send an e-mail to sigul2022@ilc.cnr.it (Subject: [SIGUL2022]).

Programme Committee (TBC)

- Mohammad A. M. Abushariah (The University of Jordan, Jordan)
- Gilles Adda (LIMSI/IMMI CNRS, France)
- Shyam S Agrawal (KIIT, India)
- Antti Arppe (University of Alberta, Canada)
- Gerard Bailly (GIPSA Lab, CNRS)
- Dorothee Beermann (NTNU, Norway)
- Laurent Besacier (LIG-IMAG, France)
- Steven Bird (Charles Darwin University, Australia)
- Federico Boschetti (CNR-ILC, Italy)
- Dewi Bryn Jones (Bangor University, UK)
- Klara Ceberio (Elhuyar, Spain)
- Emily Le Chen (University of Illinois, USA)
- Matt Coler (University of Groningen, The Netherlands)
- John Philip McCrae (National University of Ireland Galway, Ireland)
- Pradip K Das (IIT, India)
- Jelske Dijkstra (Fryske Akademy, The Netherlands)
- Adrian Doyle (University of Galway, Ireland)
- Omar Farooq (AMU, India)
- Jonas Fromseier Mortensen (Google)
- Sahar Ghannay (LIMSI, CNRS, France)
- Dafydd Gibbon (Bielefeld University, Germany)
- Jeff Good (University at Buffalo, USA)
- Atticus Harrigan (University of Alberta, Canada)
- Lars Hellan (NTNU, Norway)
- John Judge (ADAPT DCU, Ireland)
- Jordan Lachler (University of Alberta, Canada)

- Guillaume Lample (Facebook, Paris, France)
- Richard Littauer (University of Saarland, Germany)
- Teresa Lynn (ADAPT Centre, Ireland)
- Heysem Kaya (Utrecht University, The Netherlands)
- Alexey Karpov (SPIIRAS, Russian Federation)
- Irina Kipyatkova (SPIIRAS, Russia)
- Andras Kornai (Hungarian Academy of Sciences, Hungary)
- Joseph Mariani (LIMSI-CNRS, France)
- Mathieu Mangeot-Nagata (LIG-IMAG, France)
- Maite Melero (Barcelona Supercomputing Center, Spain)
- Sjur Moshagen (UiT The Arctic University of Norway, Norway)
- Satoshi Nakamura (NAIST, Japan)
- Win Pa Pa (UCS Yangon, Myanmar)
- Alexis Palmer (University of North Texas, USA)
- Hyunji "Hayley" Park (University of Illinois at Urbana-Champaign, USA)
- Delyth Prys (Bangor University, UK),
- Carlos Ramisch (Université Marseille, France)
- Sakriani Sakti (NAIST, Japan)
- Pierric Sans (Google)
- Kepa Sarasola (University of the Basque Country, Spain)
- Kevin Scannell (Saint Louis University, Missouri, USA)
- Katherine Schmirler (University of Alberta, Canada)
- Tanja Schultz (Uni-Bremen, Germany)
- Lane Schwartz (University of Illinois at Urbana-Champaign, USA)
- Lars Steinert (University of Bremen, Germany)
- Sebastian Stüker (KIT, Germany)
- Solomon Teferra Abate (Addis Ababa University, Ethiopia)
- Nick Thieberger (University of Melbourne / ARC Centre of Excellence for the Dynamics of Language,
Australia)
- Francis Tyers (Moscow Higher School of Economics, Russia)
- Trond Trosterud (Tromsø University, Norway)
- Charl Van Heerden (Saigen, South Africa)
- Daan Van Esch (Google)
- Thang Vu (Uni-Stuttgart, Germany)
- Marcelly Zanon-Boito (LIG-IMAG, France)

Acknowledgments

The SIGUL 2022 event is organized and sponsored by SIGUL¹, which serves as the Special Interest Group in Under-resourced Languages for both ELRA² and ISCA³ associations.

SIGUL 2022 is endorsed by SIGEL⁴, the ACL⁵ Special Interest Group on Endangered Languages.

Identify, Describe and Share your LRs!

When submitting a paper from the START page, authors will be asked to provide essential information about resources (in a broad sense, i.e. also technologies, standards, evaluation kits, etc.) that have been used for the work described in the paper or are a new result of your research.

Moreover, ELRA encourages all LREC authors to share the described LRs (data, tools, services, etc.) to enable their reuse and replicability of experiments (including evaluation ones).

¹ ELRA-ISCA Special Interest Group on Under-resourced Languages (<http://www.elra.info/en/sig/sigul>).

² European Language Resources Association (<http://portal.elda.org/en>).

³ International Speech Communication Association (<https://www.isca-speech.org/iscaweb>).

⁴ ACL Special Interest Group on Endangered Languages (<https://aclsigel.github.io>).

⁵ Association for Computational Linguistics (<https://www.aclweb.org>).