

Message from the Workshop Chairs

In recent years, there has been a significant upsurge in research into natural language processing (NLP) for less resourced languages. There has been a proliferation of initiatives and events such as hackathons, toolkits, special interest groups and dedicated journal issues. The topic of Lesser-used Languages has moved from a niche area of interest to a prominent place in major conferences such as LREC and ACL.

The growing interest in this research domain underscores the importance of providing a platform for collaboration and discussion focused on less-resourced languages. Equally crucial is ensuring the inclusion of the language communities and representatives of under-resourced and endangered languages to ensure that technological solutions align with their needs and aspirations, fostering open and inclusive research with substantial social impact.

The 2nd Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages (SIGUL 2023) encompasses a broad spectrum of research interests, including less-resourced, under-resourced, endangered, minority, and minoritized languages. Building on the legacy of previous events, SIGUL continues to serve as a forum for presenting cutting-edge NLP research for under-resourced languages to both academic and industry researchers. It also provides a space where researchers from diverse backgrounds and disciplines can explore new avenues of intellectual and practical development while upholding their shared commitment to preserving less-resourced languages.

To encourage interdisciplinary collaboration, this year's workshop features a joint session with the Speech and Language Technology in Education (SLaTE 2023). Additionally, given that the workshop is taking place in Dublin, we have devoted an entire session to Celtic languages.

We are also honored to have two distinguished speakers, Delyth Prys and Subhashish Panigrahi, participating in this year's event. Their lectures, along with 21 oral presentations and 5 poster presentations, covering a wide range of topics in NLP, speech, data, and general issues, are included in these proceedings. The accepted papers encompass a multitude of languages, representing 82 different languages from Europe, Asia, Africa, and the Americas. This workshop clearly demonstrates that the field of language resources and speech and natural language processing for less-resourced languages has matured and become a well-established area of study.

The SIGUL 2023 workshop is organized and sponsored by the SIGUL organization, which serves as the Special Interest Group for under-resourced languages for both the ELRA and ISCA associations. It is also endorsed by SIGEL, the ACL special interest group on endangered languages. Additionally, this year's event has received a sponsorship grant from Google Inc.