

SIGUL 2022

Workshop Programme

- All times **CEST** (Central European Summer Time)
- Room (on-site participation) and zoom link (virtual participation) to be announced

Friday, June 24, 2022

14:00-15:00 **Opening Session**

- 14:00-14:10 SIGUL 2022 Opening Talk

14:10-15:10 **Session 1: Speech**

- 14:10-14:25 (on-site) *Unsupervised Word Segmentation from Discrete Speech Units in Low-Resource Settings* –
 - Marcelly Zanon Boito, Bolaji Yusuf, Lucas Ondel, Aline Villavicencio and Laurent Besacier
- 14:25-14:40 (on-site) *An Open Source Web Reader for Under-Resourced Languages*
 - Judy Fong, Þorsteinn Daði Gunnarsson, Sunneva Þorsteinsdóttir, Gunnar Thor Örnólfsson and Jon Guðnason
- 14:40-14:55 (on-site) *Text-to-Speech for Under-Resourced Languages: Phoneme Mapping and Source Language Selection in Transfer Learning*
 - Phat Do, Matt Coler, Jelske Dijkstra and Esther Klabbers
- 14:55-15:10 (online) *ReadAlong Studio: Practical Zero-Shot Text-Speech Alignment for Indigenous Language Audiobooks*
 - Patrick Littell, Eric Joanis, Aidan Pine, Marc Tessier, David Huggins Daines and Delasie Torkornoo

15:10-16:00 **Keynote Speech**

- *Sovereignty for Under-resourced Languages*
 - Keoni Mahelona

16:00-16:30 Coffee Break

16:30-17:45 **Session 2: Data**

- 16:30-16:45 (online) *Corpus Creation for Sentiment Analysis in Code-Mixed Tulu Text*
 - Asha Hegde, Mudoor Devadas Anusha, Sharal Coelho, Hosahalli Lakshmaiah Shashirekha and Bharathi Raja Chakravarthi
- 16:45-17:00 (on-site) *Crowd-sourcing for Less-resourced Languages: Lingua Libre for Polish*
 - Mathilde Hutin and Marc Allasonnière-Tang

- 17:00-17:15 (on-site) *Tupían Language Resources: Data, Tools, Analyses*
 - Lorena Martín Rodríguez, Tatiana Merzhevich, Wellington Silva, Tiago Tresoldi, Carolina Aragon and Fabrício F. Gerardi
- 17:15-17:30 (on-site) *Quality versus Quantity: Building Catalan-English MT Resources*
 - Ona de Gibert Bonet, Ksenia Kharitonova, Blanca Calvo Figueras, Jordi Armengol-Estapé and Maite Melero
- 17:30-17:45 (online) *A Sentiment Corpus for South African Under-Resourced Languages in a Multilingual Context*
 - Ronny Mabokela and Tim Schlippe

Saturday, June 25, 2022

9:00-10:00 Session 3: MT4All

- 9:00-9:15 *General overview of unsupervised MT for under resourced languages*
 - Jordi Armengol
- 9:15-9:30 *Technical approach in MT4All*
 - Iakes Goenaga
- 9:30-9:45 *MT4All generated resources and Shared Task scope and results*
 - Ona de Gibert
- 9:45-10:00 *CUNI Submission to MT4All Shared Task*
 - Ivana Kvapilíková and Ondrej Bojar

10:00-10:30 Session 4: General Issues

- 10:00-10:15 (on-site) *Resource: Indicators on the Presence of Languages in Internet*
 - Daniel Pimienta
- 10:15-10:30 (on-site) *Language Technologies for Low Resource Languages: Sociolinguistic and Multilingual Insights*
 - A. Seza Doğruöz and Sunayana Sitaram

10:30-11:00 Coffee Break

11:00-12:45 Session 5: NLP

- 11:00-11:15 (online) *Sentiment Analysis for Hausa: Classifying Students' Comments*
 - Ochilbek Rakhmanov and Tim Schlippe
- 11:15-11:30 (online) *Nepali Encoder Transformers: An Analysis of Auto Encoding Transformer Language Models for Nepali Text Classification*
 - Utsav Maskey, Manish Bhatta, Shiva Bhatt, Sanket Dhungel and Bal Krishna Bal

- 11:30-11:45 (on-site) *CoSwID, a Code Switching Identification Method Suitable for Under-Resourced Languages*
 - Laurent Kevers
- 11:45-12:00 (online) *A Neural Network Approach to Create Minangkabau-Indonesia Bilingual Dictionary*
 - Kartika Resiandi, Yohei Murakami and Arbi Haza Nasution
- 12:00-12:15 (on-site) *Machine Translation from Standard German to Alemannic Dialects*
 - Louisa Lambrecht, Felix Schneider and Alexander Waibel
- 12:15-12:30 (on-site) *Question Answering Classification for Amharic Social Media Community Based Questions*
 - Tadesse Destaw, Seid Muhie Yimam, Abinew Ayele and Chris Biemann
- 12:30-12:45 (online) *Automatic Detection of Morphological Processes in the Yorùbá Language*
 - Tunde Adegbola

12:45-14:00 Lunch Break

14:00-15:00 [Joint SIGUL 2022 – MWE 2022 Poster Session](#)

- (SIGUL) *Evaluating Unsupervised Approaches to Morphological Segmentation for Wolastoqey*
 - Diego Bear and Paul Cook
- (SIGUL) *Baseline English and Maltese-English Classification Models for Subjectivity Detection, Sentiment Analysis, Emotion Analysis, Sarcasm Detection, and Irony Detection*
 - Keith Cortis and Brian Davis
- (SIGUL) *Building Open-source Speech Technology for Low-resource Minority Languages with SáMi as an Example – Tools, Methods and Experiments*
 - Katri Hiovain-Asikainen and Sjur Moshagen
- (SIGUL) *Investigating the Quality of Static Anchor Embeddings from Transformers for Under-Resourced Languages*
 - Pranaydeep Singh, Orphee De Clercq and Els Lefever
- (SIGUL) *Introducing YakuToolkit. Yakut Treebank and Morphological Analyzer*
 - Tatiana Merzhevich and Fabrício Ferraz Gerardi
- (SIGUL) *A Language Model for Spell Checking of Educational Texts in Kurdish (Sorani)*
 - Roshna Abdulrahman and Hossein Hassani

- *(SIGUL) SimRelUz: Similarity and Relatedness Scores as a Semantic Evaluation Dataset for Uzbek Language*
 - Ulugbek Salaev, Elmurod Kuriyozov and Carlos Gómez-Rodríguez
- *(SIGUL) ENRICH4ALL: A First Luxembourgish BERT Model for a Multilingual Chatbot*
 - Dimitra Anastasiou, Radu Ion, Valentin Badea, Olivier Pedretti, Patrick Gratz, Hoorieh Afkari, Valerie Maquil, Anders Ruge
- *(MWE) Annotating “Particles” in Multiword Expressions in te reo Māori for a Part-of-Speech Tagger*
 - Aoife Finn, Suzanne Duncan, Peter-Lucas Jones, Gianna Leoni and Keoni Mahelona
- *(MWE) Metaphor Detection for Low Resource Languages: From Zero-Shot to Few-Shot Learning in Middle High German*
 - Felix Schneider, Sven Sickert, Phillip Brandes, Sophie Marshall and Joachim Denzler
- *(MWE) Automatic Bilingual Phrase Dictionary Construction from GIZA++ Output*
 - Albina Khusainova, Vitaly Romanov and Adil Khan
- *(MWE) A BERT's Eye View: Identification of Irish Multiword Expressions Using Pre-trained Language Models*
 - Abigail Walsh, Teresa Lynn and Jennifer Foster
- *(MWE) Enhancing the PARSEME Turkish Corpus of Verbal Multiword Expressions*
 - Yagmur Ozturk, Najet Hadj Mohamed, Adam Lion-Bouton and Agata Savary
- *(MWE) German Light Verb Constructions in Business Process Models (non-archival)*
 - Kristin Kutzner and Ralf Laue

15:00-16:00 **Joint SIGUL 2022 – MWE 2022 Keynote Speech**

- *Multiword Expressions and the Low-Resource Scenario from the Perspective of a Local Oral Culture*
 - Steven Bird

16:00-16:30 Coffee Break

16:30-17:30 **Panel Discussion**

17:30-17:50 **General Discussion**

17:50-18:00 **Closing**