



# SIGUL 2023

## DAY 1st, August 18th, 2023

### 09:00-09:10 Opening Talk

09:00-09:05 SIGUL Opening Tak (SIGUL Co-Chair)

09:05:09:10 Welcome Speech (Local Organizer)

### 09:10-10:00 Keynote Speech [50min]: Subhashish Panigrahi

### 10:00-10:40 SIGUL NLP [15min presentation + 5min QA]

(on-site)	10:00-10:20	ID #14	Romanian (Romania)	Short-Cutting Manual Acquisition in Deep-Learning Deciphering of Old Documents	D. Cristea
(remote)	10:20-10:40	ID #6	Saraiki (Pakistan)	A Finite-State Morphological Analyzer for Saraiki	M. Alam, A. O'Neil, D.G. Swanson and F.M. Tyers

### 10:40-11:10 Coffee Break

### 11:10-12:30 SIGUL Resources, Pronunciation & Dialect [15min presentation + 5min QA]

(remote)	11:10-11:30	ID #31	6 Tibeto-Burman and 4 Indo-Aryan (India)	Collecting Speech Data for Endangered and Under-resourced Indian Languages	R. Kumar, M. Takhellambam, A. Gope, B Lahiri, S. Ratan, N. Mathur, S. Singh
(remote)	11:30-11:50	ID #30	Kanak (New Caledonia)	A visit to the Cliffs of Jokin: A role for phonetizers in second language pronunciation and word learning, with an example from the under-resourced languages of New Caledonia	P. Welby, B. Bigi, A. Corral, F. Wacalie, G. Wattelez
(on-site)	11:50-12:10	ID #17	Korebaju (Colombian)	An intra- and inter-dialectal study of Korebaju vowels	J.V. Rodriguez, N. Vallee, T. Chacon, C. Savariaux, S. Gerber
(on-site)	12:10-12:30	ID #12	Nepali, Lyngam, Nanasu, War, Na	From 'Snippet-lects' to Doculects and Dialects: Leveraging Neural Representations of Speech for Placing Audio Signals in a Language	S. Guillaume, G. Wisniewski, A. Michaud

### 12:30-13:30 Lunch Break

### 13:30-14:30 SIGUL POSTER SESSION [60 min]

(on-site)	13:30-14:30	ID #35	Konkani (Indo-Aryan)	Konkani Integer Phonetic Transcription System	S. Fadte, E.V. Fernades, H.H. Redkar, T. Kundaikar, R. Karmali, J.D. Pawar
(on-site)	13:30-14:30	ID #33	Maltese - Semitic (Sicilian Arabic)	The Applicability of Wav2Vec2 and Whisper for Low-Resource Maltese ASR	A. Williams, A.D. Marco, C. Borg
(on-site)	13:30-14:30	ID #22	North Sami (Finland)	Speech recognition system improvements for North Sámi Speaker-dependent and Speaker Independent tasks	A. Rouhe, M. Kurimo
(remote)	13:30-14:30	ID #18	Lule and North Sami (Finland)	Developing TTS and ASR for Lule and North Sami languages	K. Hiovain-Asikainen and J.D. Rosa
(on-site)	13:30-14:30	ID #32	English-Japanese (Japan)	VGSAAlign: Bilingual Speech Alignment of Unpaired and Untranscribed Languages using Self-Supervised Visually Grounded Speech Models	L.T. Nguyen, S. Sakti

### 14:30-16:10 SIGUL Language Technologies [15min presentation + 5min QA]

(on-site)	14:30-14:50	ID #25	Hungarian (Hungary)	What kind of multi- or cross-lingual pre-training is the most effective for a spontaneous, less-resourced ASR task?	P. Mihajlik, M.S. Kadar, G. Dobsinszki, Y. Meng, M. Kedalai, J. Linke, T. Fegyó, K. Mády
(on-site)	14:50-15:10	ID #13	Irish - Celtic (Ireland)	Towards spoken dialect identification of Irish	L. Lonergan, M. Qian, N.N. Chiaráin, C. Gobl, A.N. Chasaide
(on-site)	15:10-15:30	ID #23	Austrian-German (Austria)	Neural Speech Synthesis for Austrian Dialects with Standard German Grapheme-to-Phoneme Conversion and Dialect Embeddings	L. Gutscher, M. Pucher, V. Garcia
(remote)	15:30-15:50	ID #15	Nepali (Nepal)	Nepali Text-to-Speech Synthesis using Tacotron2 for Melspectrogram Generation	S. Khadka, R. GC, P. Paudel, R. Shah, B. Joshi
(remote)	15:50-16:10	ID #29	English-Japanese (Japan)	Low-Resource Japanese-English Speech-to-Text Translation Leveraging Speech-Text Unified-model Representation Learning	T. Tran, S. Sakti

**16:10-16:30 Coffee Break****16:30-17:30 SIGUL Language Technologies [15min presentation + 5min QA]**

(on-site)	16:30-16:50	ID #10	25 Languages (Babel)	Multilingual models with language embeddings for low-resource speech recognition	L.-M. Lam-Yee-Mui, W.B. Kheder, V.B. Le, C. Barras, J.-L. Gauvain
(on-site)	16:50-17:10	ID #19	Bambara (Mali)	Low-resource ASR-free keyword spotting using listen-and-confirm	E. Westhuizen, M. Ribeiro, J.J. Vüren, P. Hidalgo-Sanchis, T. Niesler
(remote)	17:10-17:30	ID #9	isiXhosa (South Africa)	Automatic transcription and (de)standardisation	N. Markl, E. Wallington, O. Klejch, T. Reitmaier, G. Bailev, J. Pearson, S. Robinson, M. Jones, P. Bell

**17:30-18:15 SIGUL Discussion with Panel on Language resources and technology: Inclusion, Diversity, and Sovereignty [45 min]****DAY 2nd, August 19th, 2023****09:00-10:40 SIGUL Special Session on Celtic Language - Paper Presentation [15min presentation + 5min QA]**

(on-site)	09:00-09:20	ID #16	Breton - Celtic (Britanny)	Community internally driven corpus buildings: Three examples from the Breton ecosystem	M. Joutteau
(on-site)	09:20-09:40	ID #7	Scottish Gaelic - Celtic (Scotland)	A Transformer-Based Orthographic Standardiser for Scottish Gaelic	J. Huang, B. Alex, M. Bauer, D.S. Jasin, Y. Liang, R. Thomas, W. Lamb
(on-site)	09:40-10:00	ID #11	Welsh - Celtic (Wales)	Developing Live Welsh Speech Recognition Models for a Commercial Product - a case study	P. Vangberg, L.S. Farhat, D. Jones, S. Kinahan
(on-site)	10:00-10:20	ID #26	Irish - Celtic (Ireland)	The language communities as active partners in technology provisions: the Irish ABAIR experience	A.N. Chasaide, N.N. Chiaráin, H. Berthelsen, A. Murphy, L. Lonergan, J. Sloan, C. Wendler, C. McCabe, E. Barnes, C. Gobl
(on-site)	10:20-10:40	ID #24	Irish - Celtic (Ireland)	ABAIR & ÉIST: A demonstration of speech technologies for Irish	A. Murphy, L. Lonergan, M. Qian, H. Berthelsen, C. Wendler, N.N. Chiaráin, A.N. Chasaide, C. Gobl

**10:40-11:10 Coffee Break****11:10-12:00 SIGUL Special Session on Celtic Language - Keynote Speech [50 min]: Delyth Prys****12:00-12:45 SIGUL Special Session on Celtic Language - Discussion with Panel on Role of Local Communities & Policy/Support [45 min]****12:45-14:00 Lunch Break (+ Move to Long Room Hub)****14:00-15:20 Joint SIGUL-SLaTE - Paper Presentation [15min presentation + 5min QA]**

(on-site)	14:00-14:20	(SLaTE)	Irish - Celtic (Ireland)	Mol an Óige: a phonological awareness and early literacy platform for Irish	A.N. Chasaide, E. Barnes, R. Errity, O. Mroz, O.N. Haonghusa, S.N. Chasaide, A. Giovannini, N.N. Chiaráin
(on-site)	14:20-14:40	(SLaTE)	Finnish and Finland Swedish (Finland)	New data, benchmark and baseline for L2 speaking assessment for low-resource languages	M. Kurimo, Y. Getman, E. Voskoboinik, R. Al-Ghezi, H. Kallio, M. Kuronen, A. Zansen, R. Hilden,
(on-site)	14:40-15:00	ID #27	Maori (New Zealand)	Towards Automatic Marking of Pepeha: a Formulaic Māori Language Speech	C. Watson, P. Allen, P. Keegan, K. Mahelona, P.-L. Jones
(on-site)	15:00-15:20	ID #20	Irish - Celtic (Ireland)	Geabaire, the first Irish AAC system: voice as a vehicle for change	E. Barnes, J. Cummins, R. Errity, O. Morrin, H. Berthelsen, C. Wendler, H. Husca, N.N. Chiaráin and

**15:20-15:50 Coffee Break****15:50-16:10 Joint SIGUL-SLaTE - Paper Presentation [15min presentation + 5min QA]**

(on-site)	15:50-16:10	(SLaTE)	Irish - Celtic (Ireland)	An Bat M'írialta: Stateful Development of an Irregular Verb Bot for Irish	J. Sloan, N.N. Chiaráin
(on-site)	16:10-16:30	(SLaTE)	Irish - Celtic (Ireland)	Filling the SLaTE: examining the contribution LLMs can make to Irish iCALL content generation	N.N. Chiaráin, N.R. Gunning, O. Nolan, M. Comtois
(on-site)	16:30-16:50	(SLaTE)	21 languages	ChatGPT + LARA = C-LARA	B. Bedi, B. Chiera, C. Chua, N.N. Chiaráin, M. Rayner, A. Simonson, R. Zviel-Girshin

**16:50-17:35 Joint SIGUL-SLaTE - Discussion with Panel on UL and Education [45min]****17:35-17:45 Closing and Group Photos**