

ELRA & ISCA Special Interest Group on Under Resourced Languages (SIGUL)

Candidates for the SIGUL Board 2018-2020

<http://www.elra.info/en/sig/sigul/>

<http://www.isca-speech.org/iscaweb/index.php/sigs?layout=edit&id=198>

The joint ISCA-ELRA Special Interest Group on Under-resourced Languages (SIGUL) has been created earlier in 2017. SIGUL intends to bring together a number of professionals involved in the development of language resources and technologies for under-resourced languages.

To comply with the ELRA & ISCA by-laws on SIGs, a Board has to be elected by all involved parties. Its role and duties were explained within the call for nominations.

According to these bylaws, the elected members of the Board shall consist of a chairperson, a co-chair and a secretary (see board duties in the Appendix below). All elected officers shall be members in good standing with the SIGUL organization and parent organizations (ELRA and ISCA). The term of all elected board members of the SIG shall be 2 years. They may be re-elected but consecutive appointments are limited to four consecutive terms.

Following the Call for candidates (from 22 November 2017 to 7th December 2017), we received four nominations (all valid according to by-laws).

We are now organizing an electronic vote to select three candidates out the four nominees. All experts who are/were involved in pre-SIGUL events (SLTU, CCURL, others) were compiled into a mailing list, asked to nominate the candidates for the SIGUL Board, and now asked to vote for three nominees.

We will be using the Adoodle.org voting service for this. It is a service that guarantees:

- *the list of participants is given explicitly in the sent emails, corresponding to the transparency and fairness principles of Adoodle.org*
- **what is anonymous is the action of casting a vote: no one can know who has cast a vote.**
- *once the vote is created, all its parameters are voluntarily kept fixed by the vote machine, strictly and simply applying what has been announced formally in the sent email. Therefore the vote will go to its conclusion, without allowed future modification.*

Each of you will be allowed to vote for at most 3 candidates, since there are 3 positions within the board. Your email is your identifier.

The vote starts immediately and the deadline for voting is Monday, January the 8th 23:59 CET (Central European Time, 22:59 GMT). You will be receiving an email from " no_reply@adoodle.org" with a subject like: Subject: [Adoodle.org] ELRA-ISCA Special Interest Group on Under Resourced Languages (SIGUL). Please make sure this email does not end in your spam box.

I am attaching below a short biography and a statement/manifesto that explains why the candidates want to be a member of the SIGUL board. They are in the order I received them.

A specific email will be sent from the voting site with explanations on how to vote.

Best Regards

Khalid Choukri, in charge of supervising these elections for the SIGUL committee

Laurent Besacier (Univ. Grenoble Alpes, France)

Biography

Prof. Laurent Besacier defended his PhD thesis (Univ. Avignon, France) in Computer Science in 1998 on “A parallel model for automatic speaker recognition”. Since 1999 he is an associate professor (full professor since 2009) in Computer Science at Univ. Grenoble Alpes (he was formerly at U. Joseph Fourier). From September 2005 to October 2006, he was an invited scientist at IBM Watson Research Center (NY, USA) working on Speech to Speech Translation. His research interests are mainly related to multilingual speech recognition and machine translation. Laurent Besacier has published >150 papers in conferences and journals related to speech and language processing. He supervised or co-supervised 20 PhDs and 30 Masters. He has been involved in several national and international projects as well as several evaluation campaigns. Since October 2012, Laurent Besacier is a junior member of the “Institut Universitaire de France” with a project entitled “From under-resourced languages processing to machine translation: an ecological approach”.

Research interests:

Natural Language Processing - Automatic Speech Recognition - Machine Translation - Less Resourced Languages - Machine-Assisted Language Documentation - Evaluation of NLP Systems

Why I am candidate to SIGUL board:

I am behind the creation of the workshop on Spoken Language Technologies for Under-resourced languages (SLTU) which takes place every two years since 2008. It is now a well-established venue with participants from the best research groups (for instance, Google, JHU, BBN, KIT, NICT, U. Melbourne & Waseda participated in 2016). I also participated to the new joint ELRA and ISCA Special Interest Group on Under-resourced Languages (SIGUL). If elected, I'll try my best to continue the efforts to promote under-resourced languages through organizing scientific events (conferences, special sessions, challenges), encouraging data collection efforts and help for the dissemination of corpora in low resource languages, interacting with industrial stakeholders of the field as well as with local speaking communities.

Claudia Soria (CNR-ILC, Italy)

Biography

I hold a PhD in Linguistics, and am currently employed as Researcher at Istituto di Linguistica Computazionale “A. Zampolli”, Consiglio Nazionale delle Ricerche, in Pisa, Italy (Institute of Computational Linguistics, National Research Council). I have a background in computational linguistics, with a focus on language resources in their entire life-cycle, from creation to representation to evaluation. I am one of the authors of LMF, Lexical Markup Framework, an ISO standard for the representation of computational lexicons. Since 2013 I turned my interests to minority, endangered and less-resourced languages. My current research interests revolve around use of technological means, Language Technology in particular, for the preservation and maintenance of linguistic diversity. I currently coordinate an Erasmus+ project, “The Digital Language Diversity Project”, promoting digital language diversity and access to digital media and services by speakers of lesser used languages. Other current interests are use and usability of regional/minority languages on social media; ethnolinguistic vitality of regional and minority languages of Italy; creation of lexico-conceptual resources for archiving traditional knowledge.

Why I am candidate to SIGUL board:

I promoted the creation of the ELRA SIG on less-resourced languages, which later became part of SIGUL. Over the past years, I co-chaired the CCURL workshop series, the first of which took place in 2014 in Reykjavik. I am convinced that for less-resourced languages in particular, networking, sharing of good practices and ideas is key to fast development, which is crucial for their maintenance in the digital era. If elected, I would like to continue my efforts of empowering lesser-used languages through the organization of events, dissemination of information and bottom-up collection of requirements. I would like to help making SIGUL become a big network of professionals that reaches out to speakers’ communities, in order to help them set up their own digital language planning agenda.

Sakriani Sakti (Nara Institute of Science and Technology, Japan)

Biography:

Sakriani Sakti is Assistant Professor of Graduate School of Information Science, Nara Institute of Science and Technology (NAIST), Japan. She received her B.E. degree in Informatics (cum laude) in 1999 from Bandung Institute of Technology, Indonesia. In 2000, she won DAAD-Siemens Asia 21st Century Award to study in Communication Technology, University of Ulm, Germany and earned her MSc degree in 2002. During her thesis work, she worked with DaimlerChrysler Research Center, Ulm, Germany. She then worked as a researcher at ATR Spoken Language Communication (SLC) Laboratories Japan in 2003-2009, and NICT SLC Groups Japan in 2006-2011, which established various speech recognition systems (including English, Korean, Filipino and Indonesian languages) for multilingual speech-to-speech translation. While working with ATR-NICT, Japan, she continued her study (2005-2008) at Dialog Systems Group, University of Ulm, Germany, and received her PhD degree in 2008. She actively involved in collaboration activities such as Asian Pacific Telecommunity Project (2003-2007) and various speech-to-speech translation research projects in the world, including A-STAR and U-STAR (2006-2011). In 2009-2011, she served as a visiting professor of Computer Science Department, University of Indonesia (UI), Indonesia. From 2011, she has been an assistant professor at the Augmented Human Communication Laboratory, NAIST, Japan. She also served as a visiting scientific researcher of INRIA Paris-Rocquencourt, France, in 2015-2016, under "JSPS Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation." Her research interests lie in deep learning & graphical model framework, statistical pattern recognition, zero-resourced speech technology (Her team won the first place in the Zero Resource Speech Challenge 2017 for English & Xitsonga languages), multilingual speech recognition and synthesis (including ASR systems for less-resourced languages, i.e., Javanese, Sundanese, Bataks, Bali languages), spoken language translation, social-affective dialog system, and cognitive communication.

Why I am candidate to SIGUL board:

I originally come from Indonesia, the archipelago that has more than 300 ethnic groups which speak 365 native languages or no less than 669 dialects. But 146 are already endangered, and most of them are under-resourced. Even for some languages with millions of speakers can still lack the resources needed to construct spoken and language technologies. Many of those communities do not have a chance to connect to the world. Similar phenomena also happen in many areas in Asia, Africa, and Europe. I think it would be great if we could work hand-in-hand to improve the condition, and I would be happy if I could be part of the activities. I am now a board member of Spoken Language Technologies for Under-resourced languages (SLTU) and the organizer chair of SLTU2016. I also chaired Interspeech 2017 Special Session on Digital Revolution for Under-resourced Languages (DigRevURL). I also involved in the creation of new joint ELRA and ISCA Special Interest Group on Under-resourced Languages (SIGUL). If elected, I would like to dedicate my life to support the community, not only to preserve the languages and cultures, to construct data resources, but also to accelerate the development of speech and language technologies. I would also try my best to bridge communication between countries and institutes, linguistics and engineering, researchers of natural language processing and spoken language processing, technology experts and end users, in order to establish collaborations that could provide a better future of under-resourced languages.

Professor Dorothee Beermann, PhD, Norwegian University of Science and Technology (NTNU)

Biography

I am a professor in linguistics at the Norwegian University of Science and Technology, in Trondheim, Norway.

My research over the last 15 years has had as its aim to attain a better integration of quantitative and qualitative linguistic research. I am a formal linguist familiar with most contemporary approaches to language modelling. I prefer constraint-based approaches, but have no problems applying what is on the market. I have worked extensively both with lesser described and 'mainstream' languages. In my experience work with lesser described languages requires different methodologies, and I have worked throughout the years with 'cultivated' and 'wild data' (a term coined by Joan Bresnan), and different sets of methodologies.

My approach to lesser described languages has become more and more corpus oriented over the years. I started with standard formal descriptions working on certain construction types, such as serial verb constructions in Akan and Ga, both Kwa languages of West Africa, and the Indo-Aryan language Oriya, to turn more to approaches that value the creation of digital resources as highly as the formal linguistic investigation. I have just finished a data curation project on Krio, a West African Creole language. When time allows, I engage in curating data sets from mainly African languages with the goal to make them accessible for further research. I have published over 130 papers on linguistics and NLP. As the most significant papers over the past 5 years, I should mention my work on the creation of a digital infrastructure for linguistic research as part of the Digital Humanities. I advocate Open Access to linguistic data and an approach to a digital research infrastructure where individual researchers can find it easy to use tools that allow them to create, manage and store linguistic data in an online repository (typecraft.org). I have published on this topic for example in the Springer journal Language Resources and Evaluation.

I have served in Program Committees at ACL, LREC, COLING, LTC, as well as for HPSG and LFG meetings, and been a frequent contributor at workshops directed at Lesser Resourced Languages, such as CCURL and Technologies for Indian Languages (WILDRE) at LREC, and the LRL workshops at several LTC meetings. I have arranged biannual workshops in language Resource building and Language Documentation in my own university from 2001 and till now, and co-organized workshops in the same areas at the universities in Legon and Winneba, Ghana, at Makerere University, Uganda, and in Hyderabad and Agra, India.

Why I am a candidate for the SIGUL board

The development of language resources and technologies for under-resourced languages in general has been a driving force in my research and teaching for over 10 years now. Of course I was absolutely thrilled to hear that over 300 experts have registered for the Special Interest Group on the Under- Resourced Languages established by ELRA & ISCA. It is a good development, and I would like to contribute to make it a success. My work with under-resourced languages has a strong focus on the Sub-Saharan languages of Africa, and I know that there is so much more that needs to be done to support language development in African countries.