



**HAL**  
open science

## **Introduction by the Organizers of the thematic tracks on the Achievements and Challenges (Day 2 and Day 3)**

Joseph J Mariani, Gilles Adda, Khalid Choukri, Irmgarda Kasinskaite Buddeberg, H el ene Mazo, Sakriani Sakti

### **► To cite this version:**

Joseph J Mariani, Gilles Adda, Khalid Choukri, Irmgarda Kasinskaite Buddeberg, H el ene Mazo, et al.. Introduction by the Organizers of the thematic tracks on the Achievements and Challenges (Day 2 and Day 3). Proceedings of the Language Technology for All (LT4All) conference, 2020. hal-04455045

**HAL Id: hal-04455045**

**<https://hal.science/hal-04455045v1>**

Submitted on 13 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

## **Introduction by the Organizers of the thematic tracks on the Achievements and Challenges (Day 2 and Day 3)**

*Gilles Adda, Khalid Choukri, Irmgarda Kasinskaite, H el ene Mazo, Joseph Mariani, Sakti Sakriani*

Information and knowledge in different languages is key to the achievement of sustainable development. The production of new knowledge and adequate responses to emerging challenges, such as poverty, climate change, the digital divide, uneven economic opportunities and social exclusion, will require the preservation of knowledge accumulated in local traditions and practices through the appropriate tools.

In today's world, digital literacy, access to broadband connectivity and quality content, including in local languages, are also prerequisites for the fulfillment of our human rights and fundamental freedoms, as well as for our participation in the development of societies. The denial of such rights and freedoms concerns access to knowledge and information in different languages.

Despite the policies implemented and the technological progress achieved, many language users, particularly Indigenous peoples, still experience barriers to access information online. At school, in the workplace, and in everyday life, they encounter obstacles in using the domain name system, mobile phones, computers, applications and other tools in their own languages. These difficulties are compounded by gender, disability, linguistic, economic, socio-political and digital divides.

In this context, Language Technologies (LT), greatly contribute to the promotion of linguistic diversity and multilingualism in the digital world. These are moving outside research laboratories into numerous applications in many different areas. They include keyboard technologies, spelling/grammar checkers up to speech and speaker recognition, machine translation for text and audio, speech synthesis, and spoken dialogue. They also include text and document understanding, generation and summarization, as well as sentiment and opinion analysis, answers to questions, information retrieval and knowledge access, sign languages processing, etc.

However, a small number of over the 7000 languages spoken around the world have associated Language Technologies, with various levels of quality. The majority of languages can be referred to as under-resourced or as not supported. This situation puts the users of many languages – a vast majority of Indigenous languages – in a disadvantageous situation, creating a digital divide, and placing their languages in danger of digital extinction, if not complete extinction.

For over a decade, UNESCO has been committed to promoting a new vision of inclusive knowledge societies, highlighting the importance of creation, dissemination, preservation and utilization of information and knowledge using digital technologies. This approach acknowledges the increasing role of information and knowledge in society, as powerful resources for the creation of wealth, social transformation and human development.

Therefore, UNESCO considers that it is mandatory to support linguistic diversity and multilingualism as essential pillars of fostering pluralistic, equitable, open and inclusive knowledge societies. The Organization also recognizes linguistic diversity and multilingualism as a source of enrichment for humanity and development. It also encourages its Member States to develop comprehensive language-related policies, to allocate resources and use appropriate tools in order to promote and facilitate linguistic diversity and multilingualism, including the Internet and media. In this regard, the Organization through the Recommendation concerning the "[Promotion and Use of Multilingualism and Universal Access to Cyberspace](#)", promotes the development of multilingual content and

systems, facilitation of access to networks and systems, development of public domain content and reaffirmation of equitable balance between the interests of rights holders and the public interest.

Furthermore, UNESCO as a lead UN agency for the organization of the 2019 International Year of Indigenous Languages has worked to raise awareness on the critical status of many indigenous languages around the world, and to mobilize necessary resources for an immediate action as well as the implementation of the Action Plan for organizing the 2019 International Year of Indigenous Languages.

Based on the Roadmap towards [UNESCO's World Atlas of Languages](#), the Organization is already developing the online platform that will contain data on languages, policies, regulations, technical recommendations and best practices in this field. It is expected that a new globally accessible and open online platform will be used for monitoring and promotion of the world's languages online, strengthen cooperation and knowledge sharing. This will be by using open and inclusive technological solutions among international, regional and national language institutions, language users and other public and private partners.

## **Key messages**

The development of Language Technologies provides opportunities to improve the free flow of ideas by word and image in different languages and should leave no one behind regardless of the users' age, gender, abilities, language or location.

All language users should have access and be able to use Language Technologies, to provide and receive appropriate content and services in their own language.

Equipping language users with the necessary tools to benefit from the latest digital developments requires the joint and long-term efforts of all stakeholders, including governments, language users, in particular Indigenous peoples, academia, civil society and the private sector.

## **Major objectives, expected outcomes and outputs**

Within the framework of the 2019 International Year of Indigenous Languages, the 3-day International Conference looked at the evolving environment of linguistic diversity and multilingualism which play an important role in our societies. The event also aims to identify recommendations on how to harness technology for the preservation, support and promotion of languages, including lesser-used and Indigenous languages, as well as on how to increase and facilitate communication between language users. The Conference will conclude its work with an Outcome Document in order to set out the direction for future global actions in the area of Language Technologies.

With a view to provide access to information and knowledge to all language users and facilitate their inclusion and participation in building sustainable knowledge societies, UNESCO, in close cooperation with the Government of the Khanty-Mansiysk Autonomous Okrug-Ugra (Russian Federation), the European Language Resources Association ([ELRA](#)) and its Special Interest Group on Under-resourced languages (SIGUL), and in partnership with UNESCO Intergovernmental Information for All Programme (IFAP) and the Interregional Library Cooperation Centre, as well as with support of other public organizations and sponsors, has organized this conference with the title "**Language Technologies for All: Enabling Linguistic Diversity and Multilingualism Worldwide**" at its Headquarters in Paris, France from 4 to 6 December 2019.

The major objective of the event is to promote the human rights and fundamental freedoms of all language users to access and create information and knowledge in language they best understand and to encourage all relevant stakeholders to take concrete measures for the promotion of linguistic diversity and truly multilingual internet and Language Technologies, with special focus on Indigenous languages within the context of the international normative instruments and international cooperation mechanisms such as the 2019 International Year of Indigenous Languages ([IYIL2019](#)).

**The specific objectives are to:**

- Identify existing challenges and barriers for language users to access and create information and knowledge in different languages,
- Explore the relationship between technologies and languages from a scientific, technical, cultural, linguistic, economic and political perspective,
- Look at the socio-cultural impact of technologies on languages and their users, with special focus on lesser-used, minority and Indigenous languages,
- Raise awareness of the significance and complexity of linguistic diversity and multilingualism, and the sense of urgency for taking concrete measures and actions by all relevant stakeholders,
- Share good practices on existing technological solutions to tackle the growing linguistic divide between users of different languages.

**The expected outputs:**

- Concrete recommendations to improve the status of linguistic diversity and multilingualism with regards to the role of language technologies in increasing access to multilingual information and knowledge,
- Research and identification of good practices and technical solutions for the promotion of linguistic diversity and multilingualism
- More public-private stakeholders engaged in the production of multilingual content, the establishment of enabling environments and further deployment of appropriate Language Technologies for the promotion of linguistic diversity and multilingualism,

**The expected Outcome:**

- Long-lasting commitment of all stakeholders to the development of new Language Technologies, as an effective and necessary means for promoting linguistic diversity and multilingualism.

New technology such as Artificial intelligence is used in providing translations between languages, documentation of languages and learning languages through natural language and voice interface platforms. While technology is advancing in artificial intelligence, some languages are disadvantaged by the economic models prevalent in such emerging technologies.

The International conference has therefore also taken stock of the technological state-of-the-art in language technologies, including Artificial Intelligence methods such as Machine Learning, including Deep Learning, and its need for linguistic data.

The 3-day conference was composed of keynote talks, oral and poster presentations, panel discussions as well as reports, demonstration of systems and innovative applications, and experiences gained in the deployment of technologies in promoting linguistic diversity and multilingualism. Innovative solutions were presented at the exhibition specially installed during the event.

The event will draw conclusions and recommendations in an **Outcome Document** resulting from the presentations, panel discussions and written contributions.

The conference gathered UNESCO Member States, international bodies, regional governments and administrations, academia, language technology researchers, linguists, industry, indigenous peoples' and language policy and decision makers from around the world.

The attendance at the conference was on invitation and the audience was targeted to a maximum of 400 people.

## **Committees**

The International Conference was organized by UNESCO, Government of the Khanty-Mansi Autonomous Area – Ugra and the European Language Resources Association (ELRA) and its Special Interest Group on Under-resourced languages (SIGUL), in partnership with the UNESCO Information for All Programme (IFAP) and Interregional Library Cooperation Centre.

The Conference was conducted by a **Coordination Committee** that is chaired by the Assistant Director-General for Communication and Information of UNESCO. Other members of the Organizing Committee include a representative from the:

- Government of the Khanty-Mansi Autonomous Area – Ugra (Mr Alexey Shipilov, First Deputy Governor of the Khanty-Mansiysk Autonomous Area – Ugra),
- European Language Resources Association – (Mr Khalid Choukri, General Secretary, European Language Resources Association ELRA, Mr Joseph Mariani, Honorary President of ELRA, and Ms Sakriani Sakti, Secretary, Joint Special Interest Group on Under-resourced languages, SIGUL),
- UNESCO Intergovernmental Information for All Programme (Ms Dorothy Gordon, Chair and Mr Evgeny Kuzmin, Vice-Chair),
- Steering Committee for the organization of the 2019 IYIL2019 (Ms Aili Keskitalo, Indigenous Co-chair and Chair of the Ad-hoc group on Language Technologies),
- UNESCO Secretariat (Ms Irmgarda Kasinskaite-Buddeberg and Mr Jaco du Toit).

The Coordination Committee is supported by sub-committees for Day 2 and 3:

- International Honorary Advisory Committee,
- Program Committee,
- Organizing Committees,
- Outcome Document Drafting Committee.

## **International Honorary Advisory for the thematic tracks on the Achievements and Challenges:**

Ahmed Boukous, Rector, Institut Royal de la Culture Amazighe (IRCAM), Morocco  
António Branco, Professor, University of Lisbon & President of the European Language Resources Association (ELRA), Portugal  
Nicoletta Calzolari, Research Fellow, Institute for Computational Linguistics "A. Zampolli - CNR & Honorary President of the European Language Resources Association (ELRA) and LREC General Chair, Italy  
Myrna Cunningham, President of the Fund for development of Indigenous Peoples of Latin America and the Caribbean (FILAC) & Former Chair, UN Permanent Forum on Indigenous Issue, Nicaragua  
Lang Fafa Dampha, Executive Secretary, African Academy of Languages (ACALAN), African Union Commission, Mali  
Attie de Lange, Director, South African Centre for Digital Language Resources (SADiLaR), South Africa  
Véronique Delvaux, FNRS Research Associate, University of Mons & President, Association Francophone de la Communication Parlée (AFCP), Belgium  
Jill Evans, European MP for Wales, UK  
Vigdís Finnbogadóttir, Former President of the Republic of Iceland & UNESCO Goodwill Ambassador for Language Diversity, the Vigdís Finnbogadóttir Institute of foreign Languages, University of Iceland, Iceland  
Martine Garnier-Rizet, Head of Digital Technology and Mathematics Department, ANR, France  
Dorothy Gordon, Chair, Intergovernmental Council, UNESCO Information for All Programme (IFAP), Ghana  
John Hansen, Professor, University of Texas at Dallas & President, International Speech Communication Association (ISCA), USA  
Aili Keskitalo, President, Sámi Parliament, Norway  
Mark Liberman, Professor, University of Pennsylvania & Director, Linguistic Data Consortium (LDC), USA  
Joseph Lo Bianco, Former President, Australian Academy of the Humanities & Professor of Language and Literacy Education, University of Melbourne, Australia  
Sixto Molina, Head of Secretariat, European Charter for Regional or Minority Languages DGII & Directorate of Anti-Discrimination, Council of Europe, France  
Makoto Nagao, Professor Emeritus, Kyoto University, Member of the Japan Academy, Japan  
Satoshi Nakamura, Professor, Nara Institute of Science and Technology (NAIST) & General Convenor, Cocosda & Oriental Cocosda, Japan  
Dina Ocampo, Former undersecretary, DepED (Department of Education) & Professor, College of Education, University of the Philippines, the Philippines  
Patrick Paroubek, Senior Research Engineer, LIMSI-CNRS & President, Association pour le Traitement Automatique des Langues (ATALA), France  
Adama Samassekou, Founding Executive Secretary, AU/ACALAN, President, WSIS PrepCom for the Geneva Phase, President, MAAYA the World Network for Linguistic Diversity & Former Minister of Education, Mali  
Algirdas Saugardas, Former European Deputy & former Minister of Foreign Affairs, Lithuania  
Paul de Sinety, Délégué Général à la Langue Française et aux Langues de France, DGLFLF, France  
Ming Zhou, Principal researcher, Microsoft Research Asia & President, Association of Computational Linguistics (ACL), China

Chengqing Zong, Research Fellow, Chinese Academy of Sciences & President, Asian Federation of Natural Language Processing (AFNLP), China

**Program Committee for the thematic tracks on the Achievements and Challenges:**

Tunde Adegbola, Executive Director, African Languages Technology Initiative (Alt-i), Nigeria  
Mirna Adriani, Dean, Computer Science Faculty, Universitas Indonesia, Indonesia  
Shyam Agrawal, KIIT Group of Colleges, Gurgaon, India  
Ahmed Ali, Qatar Computing Research Institute, Hamad Bin Khalifa University, Qatar  
Fadoua Ataa Allah, Institut Royal de la Culture Amazighe (IRCAM), Morocco  
Antti Arppe, Alberta Language Technology Laboratory (ALTLab), University of Alberta, Canada  
Kalika Bali, Microsoft Research, India  
Nicholas Barla, Indigenous Peoples Forum Odisha (IPFO), India  
Dorothee Beermann, Norwegian University of Science Technology (NTNU), Norway, Polytext AS, Norway  
Martin Benjamin, Kamusi Project International, Switzerland  
Laurent Besacier, Laboratoire d'Informatique de Grenoble (LIG), France  
Brigitte Bigi, Laboratoire Parole et Langage (LPL), France  
Steven Bird, Charles Darwin University, Australia  
Sonia Bosch, University of South Africa, South Africa  
Karim Bouzoubaa, Department of computer science, Mohammadia School of Engineers, Mohammed V, University in Rabat, Morocco  
Alena Butryna, Google Research, USA  
Chris Cieri, Executive Director, Linguistic Data Consortium (LDC), UPenn, USA  
Franciska de Jong, Executive Director, CLARIN European Research Infrastructure for Language Resources and Technology, The Netherlands  
Emmanuel Dupoux, EHESS, France  
Salwa el Ramly, President, Egyptian Society of Language Engineering (ESOLE), Egypt  
Vicent Fenollar i Sastre, Policy and Outreach Manager, Network to Promote Linguistic Diversity (NPLD), Belgium  
Colleen Fitzgerald, Former Director for Documenting Endangered Languages (DEL-NSF), University of Texas at Arlington, USA  
Mikel Forcada, Universitat d'Alacant, Spain  
Philippe Gelin, Head of Sector Multilingualism, DG/CONNECT - European Commission, Belgium  
Paul Geraghty, University of the South Pacific ,Suva, Fiji  
Dafydd Gibbon, Bielefeld University, Germany & Jinan University Guangzhou, China  
Lea Gimpel, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, Germany  
Thibault Grouas, Délégation Générale à la Langue Française et aux Langues de France (DGLFLF), France  
Ximena Gutierrez-Vasques, Universidad Nacional Autonoma de Mexico, Mexico  
Dieter Halwachs, Plurilingualism Research Unit, University of Graz, Austria  
Thomas Hanke, University of Hamburg, Germany  
Mary Harper, USA  
Mark Hasegawa-Johnson, University of Illinois, USA  
Auður Hauksdóttir, The Vigdís Finnbogadóttir Institute of Foreign Languages, University of Iceland, Iceland  
Julia Hirschberg, Columbia University NYC, USA  
Misako Ito, Regional Adviser for Communication and Information, UNESCO Bangkok,

Thailand

Mustafa Jarrar, Computer Science Dept Birzeit University, Palestine

Kristiina Jokinen, University of Helsinki, Finland

Alexey Karpov, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Russia

Sanjeev Khudanpur, Johns Hopkins University, USA

Sabine Kirchmeier, Vice-President, EFNIL (European federation of National Institutions for Languages), and Danish Language Council, Denmark

Kate Knill, Machine Intelligence Laboratory, Cambridge University, UK

Andras Kornai, Computer Science Research Institute, Hungarian Academy of Sciences (MTA SZTAKI), Hungary

Simon Krek, Jožef Stefan Institute, Slovenia

Terry Langendoen, University of Arizona, USA

Eirik Larsen, Political Advisor, Sámi Parliament, Norway

Gina-Anne Levow, University of Washington, USA

Haizhou Li, National University of Singapore (NUS), Singapore

Aijun Li, Institute of Linguistics Chinese Academy of Social Sciences, China

Lydia Liu, Institute for Comparative Literature and Society (ICLS), Columbia University, USA

Sjur Nørstebø Moshagen, Head of Sámi language technology application development, UiT The Arctic University of Norway, Norway

Shri Narayanan, University of Southern California, USA

Girish Nath Jha, Jawaharlal Nehru University (JNU), India

Graham Neubig, Carnegie Mellon University, USA

Ailbhe Ní Chasaide, Trinity College Dublin, Ireland

Nathaniel Oco, The Philippines

Nicholas Ostler, Chair, Foundation for Endangered Languages (FEL), UK

Win Pa Pa, University of Computer Studies Yangon (UCSY), Myanmar

Cecilia Piaggio, Founder, LatinoAmericaHabla, Argentina

Thierry Poibeau, Ecole Normale Supérieure Paris, France

Delyth Prys, Bangor University, Wales, UK

Georg Rehm, DFKI Berlin, Germany

Michael Rießler, University of Eastern Finland, Finland

Clara Rivera, Google Research, UK

Mike Rosner, Department of Artificial Intelligence, University of Malta, Malta

Priyankoo Sarmah, Indian Institute of Technology Guwahati, India

Yusuf Sawaki, Center for Endangered Languages Documentation (CELD), Universitas Negeri Papua, Indonesia

Odette Scharenborg, Delft University of Technology, The Netherlands

Holger Schwenk, Facebook AI Research, France

Mandana Seyfeddinipur, SOAS Endangered Languages Documentation Programme, University of London, UK

Tanja Schultz, Cognitive Systems Lab (CSL), Universität Bremen, Germany

Claudia Soria, Istituto di Linguistica Computazionale "A. Zampolli"- CNR, Italy

Virach Sornlertlamvanich, SIIT Thammasat University, Thailand & Musashino University, Japan

Juan Steyn, South African Centre for Digital Language Resources (SADiLaR), South Africa

Sebastian Stüker, Karlsruhe Institute of Technology, Germany

Akatsuki Takahashi, Advisor for Culture at UNESCO Office for the Pacific States, Samoa

Nick Thieberger, The Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC), University of Melbourne The ARC Centre of Excellence for the Dynamics of



Language, Australia

Charu Bikash Tripura, Regional Capacity Building Coordinator Asia Indigenous Peoples Pact (AIPP), Thailand

Jan Trmal, Center for Language and Speech Processing (CLSP), Johns Hopkins University, USA

Alexey V. Tsykarev, Chair, Center of Support of Indigenous People and Civic Diplomacy "Young Karelia", Former Chair, UN Expert Mechanism on the Rights of Indigenous People & Incoming Expert Member, UN Permanent Forum on Indigenous Issues, Russia

Francis Tyers, Indiana University, USA

Charl van Heerden, Saigen, South Africa

Zygmunt Vetulani, Adam Mickiewicz University in Poznan, Poland

Lining Wang, Center for the Protection and Research of Language Resources of China Beijing Language and Culture University (BLCU), China

Daniel Willett, Amazon Alexa R&D, Aachen, Germany

Masahiro Yamada, National Institute for Japanese Language and Linguistics (NINJAL), Japan

### **Organizing Committee for the thematic tracks on the Achievements and Challenges:**

Gilles Adda, LIMSI-CNRS, France

Khalid Choukri, ELRA, France

Irmgarda Kasinskaite-Buddeberg, UNESCO, France

Joseph Mariani, LIMSI-CNRS, France

Hélène Mazo, ELRA, France

Sakriani Sakti, NAIST/RIKEN AIP/SIGUL, Japan

### **Format and structure**

The International Conference programme of three days, consisted of plenary oral and poster sessions, introductory presentations and panel discussions in each session, as well as the exhibition on innovative language technologies and supporting events (such as welcome receptions).

An Outcome Document will capture the main debates and discussions and will seek to draw concrete recommendations and solutions for informed decision making regarding international cooperation, research and development. The document will be globally distributed among various stakeholders and integrated into new follow-up activities, including to the 2019 International Year of Indigenous languages.

### **Language Technology exhibition**

The exhibition showcased Language Technology solutions by both public and private partners, with relevance to the theme of the conference.

## **Program Overview:**

### **Day 1. CONTEXT. Multilingualism for building knowledge societies**

Opening Session

Keynote address: Adama Samassekou, H.E. Ibrahim Albalawi

Moderator: Evgeny Kuzmin

Oral Session 1: The multilingual landscape and Indigenous Languages

Moderator: Gilvan Muller de Oliveira

Oral Session 2: Languages and Communication in the 21st century

Moderator: Alisher Ikramov

Oral Session 3: Language policies and globalization: challenges and opportunities

Moderator: Anuradha Kanniganti

Oral Session 4: Multilingualism in education and research

Moderator: Coetzee Bester

Welcome reception

### **Day 2. ACHIEVEMENTS. Applying language technologies for linguistic diversity and multilingualism**

Opening Session

Keynote: Daan van Esch

Moderators: Marko Grobelnik, Heather Souther.

Rapporteur: Eugenia Urrere

Poster Session 1: Pacific Languages

Moderators: Steven Bird, Apolonia Tamata, Isabella Shields

Oral Session 5: Innovative aspects related to applications of Language Technologies in various areas, products and services

Moderators: Febe de Wet, Hermann Ney.

Rapporteur: Ethel Ong

Poster Session 2: European and Arctic Languages

Moderators: Teresa Lynn, Daniil Kocharov.

Oral Session 6: Scientific aspects related to the state of the art in Language Technologies, for spoken, written and sign languages

Moderators: Volker Steinbiss, Ximena Gutierrez-Vasques.

Rapporteur: Dessi Puji Lestari.

Poster Session 3: Latin America and the Caribbean Languages

Moderators: Francisco Cláudio Sampaio de Menezes, Marco Antonio Martínez Pérez, Anuschka van 't Hooft

Oral Session 7: Infrastructural aspects

Moderators: Mark Liberman, Amanda Harris.

Rapporteur: Alexey Karpov

Social event

Invited Talks Tunde Adegbola,  
Rory O'Connor

Moderators: Sakriani Sakti, Khalid Choukri, Joseph Mariani.

### **Day 3. CHALLENGES. Addressing the digital divide and multilingualism**

Keynote: Lorna Williams

Moderators: Justus Roux, Sanjeev Khudanpur.

Rapporteur: Priyankoo Sarmah.

Oral Session 8: Minority and Indigenous Languages

Moderators: Yoshinori Sagisaka, Sonja Bosch.

Rapporteur: Claudia Soria

Poster Session 4: African Languages

Moderators: Dorothy Beermann, Joseph Nkonga, Audace Niyonkuru

Oral Session 9: Activities for language preservation, reclamation, and enhancement

Moderators: Cyntia Montaña, Mikel Lorenzo Forcada Zubizarreta.

Rapporteur: Satoshi Tamura

Poster Session 5: Asian Languages

Moderators: Budi Irmawati, Netra Mani Rai

Oral Session 10: Scientific aspects related to handling language diversity

Moderators: Laurent Besacier, Gilles Adda.

Rapporteur: Nathaniel Oco

Poster Session 6: USA and Canada Languages

Moderators: Chris Cieri, Francis Tyers, Jan Trmal

Oral Session 11: Developing Language Technologies for All : Best Practices

Moderator: Delyth Prys.

Rapporteur: Emiliana Cruz

Oral Session 12: The Future of Language Technologies for All: Outcome Document and Recommendations and Closing Remarks

## **Opening speech from Dr Moez Chakchouk, Assistant Director-General for Communication and Information of UNESCO (5th of December 2019)**

Good morning,

- Excellencies
- Distinguished guests
- High Representatives of China, France, New Zealand, Norway
- Ladies and gentlemen

It is my great pleasure to address you again, on the second day of this International Conference. Thanks to all of you for being here today, despite current transportation challenges.

I would also like to repeat our gratitude to:

- the Government of the Khanty-Mansiysk Autonomous Okrug-Ugra (Russian Federation);
- the European Language Resources Association (ELRA) and its Special Interest Group on Under-resourced languages (SIGUL);
- UNESCO Intergovernmental Information for All Programme (IFAP);
- the Interregional Library Cooperation Centre;
- as well as additional public and private organizations and our sponsors.

Your invaluable institutional and financial support has enabled this timely event.

This conference is organized in the context of the 2019 International Year of Indigenous Languages, proclaimed by a UNGA resolution that appointed UNESCO as lead UN Agency for its implementation. This international year has aimed to raise awareness about the critical status of indigenous languages worldwide, and the need to mobilize stakeholders and resources to promote, revitalize and support them.

In an increasingly digitalized world, such outcomes can only be reached if concrete efforts are made to close the existing digital gap between dominant, official languages and minority, lesser-used and indigenous ones.

Throughout this international year, nearly nine-hundred (900) events have been taking place all over the world. More than eighty-one thousand (81,000) individuals have joined the global online community of the dedicated website, as well as on social media.

Throughout 2019, UNESCO has carried out a series of regional consultations to identify recommendations for future actions. These have been integrated in the Global Strategic Outcome Document for the International Year. This Document includes a key conclusion on language technologies' role in influencing development and contributing to intergenerational transmission of indigenous languages. Now, as the International Year is drawing to a close, this international conference at UNESCO represents a unique occasion reflect on future actions in the area of language technologies.

For our part, since several decades UNESCO has recognized linguistic diversity and multilingualism as a source of enrichment for humanity and development. In 2003, our Member States' agreed on the only normative instrument that currently exists at the UN level for the promotion of linguistic diversity. It is called the *Recommendation concerning the Promotion and Use of Multilingualism and Universal Access to Cyberspace*. It is a call for governments to develop comprehensive language-related policies, and to allocate resources and use appropriate tools, including the Internet and media.

There is indeed new potential today, thanks to technological advance. The engagement of major tech and IT companies in the promotion of linguistic diversity and multilingualism is vital for the future of inclusive communication that does not leave indigenous languages behind.

Let us therefore also keep awareness that relatively few language technologies have been developed in lesser-used, minority, and indigenous languages, partly due to the unavailability of rich datasets in these languages. Additionally, speakers of these languages do not often have the means and skills to develop strategies to promote their own languages online, or their integration into educational programs and cultural industries.

To understand and map the imbalances, in 2020 UNESCO will launch an online platform for the World Atlas of Languages, as a repository for linguistic diversity and multilingualism. It will be based on a global data collection initiated by UNESCO in close cooperation with the UNESCO Institute of Statistics (UIS), and on the work of leading experts. We as UNESCO will also next year publish a World Report on the status of languages, which will include new information and detailed data on human languages, including official and unofficial, spoken as well as sign languages. We hope these will be of use to you going ahead.

Meantime, I am confident that you, the experts and professionals, gathered during these days will devise strategies to address the many challenges to ensure the development and mainstreaming of language technologies in minority, lesser-used languages, as well as the main ones. Such multi-stakeholder partnership is one of the key pillars on which the Action Plan of the International Year is based.

This spirit of co-operation should inform new research directions and business strategies to help reduce the cost for developing language technologies, and to encourage an increase in knowledge-sharing practices. This will be the role of the forward-looking Outcome Document of your deliberations. Your recommendations in this Outcome Document will help shape cooperation, research and development.

Your recommendations will also be significant in the light of the proposed declaration of a decade of indigenous languages between 2022 – 2032, which will be decided by the United Nations General Assembly (UNGA) later this month. In other words, the future is there for us to make.

I thank you for your attention and wish you fruitful deliberations.

## Speech from Dr Khalid Choukri, ELRA Secretary General (5th of December 2019)

Excellencies

Ladies and gentlemen

Dear Colleagues

I am very honored to welcome you all on behalf of the European Language Resources Association. Being a member of an indigenous community myself makes this conference a very special event.

I am very honored to be here today to share with you both our hopes, our plans and our objectives, in organizing and supporting this first **LANGUAGE TECHNOLOGY FOR ALL** conference. We are confident that we are setting a major sustainable milestone that we will develop for the years to come.

I am responsible for the operations of the European language Resources association and would like to take this opportunity to share with you some of our visions.

The European language resource association (ELRA) is a non-for profit organization based in Europe and acting globally, set up in 1995 (so this is our 25<sup>th</sup> anniversary) with a primary mission to support the development of language technologies by making available and shareable language resources.

ELRA is a membership-based organization and we are very proud to have among our over 50 members some of the key players from industry and academia but also over 1000 individual members from all areas of language technologies, worldwide.

When analyzing the main stream of Language Technologies, we see that today's key paradigm, (and we will review the situation over the next two days), is the data driven paradigm. Whether we talk about Machine Learning, Deep Learning or simply Artificial Intelligence, the main fuel remains data, and we often state that the best data is more data.

ELRA has managed to secure more than 1400 resources covering. all modalities from Audio/visual, to text, terminology, lexicographical data, sign language resources, data for OCR, and many more modalities, to ensure that several language technologies can be developed.

Unfortunately, when I look through the language resources catalog but also other data centers, I see that the number of languages we serve is very low; not more than 100 and not all of them with enough data for the development of real applications. This shows the efforts needed to cover even very very partially the needs of the 7000 human languages.

To ensure that the identified resources are usable, we pay serious attention to the ethical aspects as well as legal frameworks when collecting and processing such data.

Our expertise in clearing all Intellectual Property Rights (IPR) helps our partners to have a legal access to the resources they need but also ensures that the right holders get the right credit and acknowledgment, in particular when resources are shared free of charge. Addressing these IPR issues at the early stages of data collection and production foster the use and reuse of language resources.

In addition to our mission of identifying language resources, negotiating the distribution agreements and licenses, ELRA also produces resources on demand both for research and commercial applications, very often in partnership with our local colleagues.

To illustrate this production process I would like to highlight the role of public policies and exemplify this through two cases.

Almost 20 years ago, the European Union decided to co-fund the development of speech resources for most of the European Union languages. We established a consortium that collected data for telephony, automotive, and entertainment applications, among others, and extended the language coverage to Latin America, North Africa, and Asia. These resources are very likely an important trigger of some of the today's speech interaction and dictation applications we see on our smartphones.

The second scenario is also a European Commission funded project initiated a few years ago in the framework of a European regulation called "Public sector information directive". Under this directive, all public data have to be made available both for research and industry. An important part of the data that we collected with our partners of the European Language Resource Consortium consists of parallel corpora with aligned multilingual texts that have been used to improve the performance of the Neural-based Machine Translation services used by the European public bodies.

Why I am quoting these cases.

The main reason is to highlight the need for public policies but also for public-private partnerships to produce the required resources. The production processes have gained a tremendous efficiency and cost-effectiveness today compared to 20 years ago, which can be put forward as an argument for investments in Language Resources.

At the start of ELRA, working together with another European network, ELSNET, we introduced the concept of Basic Language Resource Kit, a Language Resource and Language Technology package that would comprise a minimal set of resources that can encourage researchers to tackle the needs of any language. ELRA advocates for open source policy, in particular for the under resourced Languages and the indigenous languages.

Another important achievement of ELRA is the set-up of a major conference, language resources and Evaluation conference (LREC) that focuses on language resources and evaluation of technologies. LREC, set up in 1998, takes place every two years, next one is in n May 2020 here in France Marseille.

LREC attracts over 1300 participants that discuss all issues related to language resources I mentioned before. ELRA brings together experts from all language technology fields working on speech interaction, spoken dialogue speech to speech translation machine translation sentiment analysis sign language issues emotion and affective computing, optical character recognition text analysis lexicography terminology etc.

But in addition to the main conference, LREC makes room to a large number of satellite workshops and tutorials which are small events (with 20 to 50 experts getting together) to discuss various topics from sign language issues to global lexicography to Arabic, Semitic, Indian, African languages etc.

During LREC we also cover topics that we will address today and tomorrow with sessions dedicated to infrastructures, strategic roadmapping etc.



ELRA is also partnering with other organizations to create more synergies. To this end, ELRA and ISCA (international speech Communication association) have established a joint special interest group on Under-Resourced Languages. The core mission of this Special interest Group is to support linguistic diversity through language technologies and promotes research activities for languages that don't benefit of substantial resources in the Digital world.

The special interest group organizes workshops and tutorials to develop its mission. With over 300 members from 34 countries we are looking forward to high-value achievements in the near future.

ELRA feels very proud to see the hundreds of languages that have now access to technology and measure how far we have come. But also foresees the road ahead, a long and difficult one, that require the involvement of all of us: citizens from indigenous populations, civil society/NGO, as well as government and policy makers. Under a clear ethical statement that indigenous communities should have the capability to define their own needs, expectations, and requirements.

We are very proud to join forces with UNESCO and Khanty-Mansi Region to bring in more experts and language activists to help us draw attention to the requirements and the expectations of indigenous languages that don't have all the resources needed by today's technology.

I wish you, I wish us, a very fruitful language technology for all conference

Khalid Choukri ELRA Secretary General and ELDA CEO