Organized and supported by

United Nations
Educational, Scientific and
Cultural Organization

2019 International Year
of Indigenous Languages

Government of
the Khanty-Mansiysk
Autonomous Okrug – Ugra
of Russian Federation

EUROPEAN LANGUAGE RESOURCES ASSOCIATION

International Conference
Language Technologies for All (LT4All):
Enabling Linguistic Diversity and Multilingualism Worldwide

in the framework of
the 2019 International Year of Indigenous Languages

4 - 6 December 2019
UNESCO Headquarters, Paris, France

Conference programme at a glance

In partnership with

United Nations
Educational, Scientific and
Cultural Organization

IFAP
Information for All
Programme

INTERREGIONAL LIBRARY
COOPERATION CENTRE

SIG-UL
Special Interest Group: Under-resourced Languages
An ELRA-ISCA Joint SIG

Programme
5 December 2019 (Room II)

Day 2. ACHIEVEMENT
Applying language technologies for linguistic diversity and multilingualism

(Room II and Poster Area: Hall Segur)

08:00-09:00 Registration

09:00-10:15 Opening Session

Moderator: Mr Moez Chakchouk, Assistant Director-General for Communication and Information, UNESCO

Mr Moez Chakchouk	Assistant Director-General for Communication and Information	UNESCO	
Ms Anne Karin Olli	Vice Minister	National Government, Ministry of Local Government and Modernisation	Norway
Ms Dorothy Gordon	Chair	UNESCO Intergovernmental Information for All Programme (IFAP)	Ghana
Ms Aiolupotea Sina Aiono	Deputy Chief Executive, Regional Partnerships	New Zealand Ministry for Pacific Peoples	New Zealand
Mr Khalid Choukri	Secretary General	European Language Resources Association (ELRA) and Special Interest Group: Under-resourced languages (SIGUL)	France
Ms Lixin Tian	Director-General	Department of Language Information Management, Ministry of Education	The People’s Republic of China

Introduction of the LT4ALL: Language Technologies for All by the co-chairs of LT4All Day 2-3

10:15-10:30	Ms Sakriani Sakti	Associate Professor, Nara Institute of Science and Technology, Japan Secretary of SIGUL, ELRA-ISCA Special Interest Group of Under-Resourced Languages
	Mr Joseph Mariani	Senior Researcher, LIMSI-CNRS, France Honorary President of ELRA, France
	Mr Khalid Choukri	Secretary-General, European Language Resources Association / ELRA CEO

10:30-11:00 Session Keynote 1

Moderators:

Mr Marko Grobelnik	Researcher, Jozef Stefan Institute, Slovenia
Mr Heather Souther	Program Coordinator, Prairies to Woodlands, Indigenous Language Revitalization Circle, Canada

***Rapporteur*:**

Ms Eugenia Urrere	Director, Latinoamerica Habla, Argentina

Mr Daan van Esch Senior Technical Program Manager, Google, Inc., USA
Building Language Technologies for Everyone

11:00-11:30 Poster Session P1: Pacific Languages
Coffee Break - Poster Area Hall Segur

Moderators/Rapporteurs:

Mr Steven Bird	Professor, Charles Darwin University, Australia
Ms Apolonia Tamata	Senior Culture & Heritage Specialist, iTaukei Trust Fund, Fiji
Ms Isabella Shields	Research Assistant, University of Auckland, New Zealand

P.1.1 Ms Ana Krajinovic Humboldt-Universität zu Berlin/The University of Melbourne Germany/Australia
Ms Ana Krajinovic, Ms Rosey Billington, Mr Lionel Emil, Mr Gray Kaltap̃au and Mr Nick Thieberger
Building capacity for community-led documentation in Erakor, Vanuatu

P.1.2 Ms Amanda Harris University of Sydney, Australia
Ms Amanda Harris and Mr Nick Thieberger
PARADISEC (Pacific and Regional Archive for Digital Sources in Endangered Cultures)

P.1.3 Mr Paul Nelson SIL International, USA
Mr Paul Nelson
Bloom Books

P.1.4 Ms Isabella Shields University of Auckland, New Zealand
Ms Isabella Shields, Ms Catherine Watson, Mr Peter Keegan, Ms Rebekah Berriman and Ms Jesin James
Creating a Synthetic Te Reo Māori Voice

11:30-13:00 Session 5: Innovative aspects related to applications of Language Technologies in various areas, products and services

Moderators:

Ms Febe de Wet	Research Associate, Stellenbosch University, South Africa
Mr Hermann Ney	Professor, RWTH Aachen University, Germany

Rapporteur:

Ms Ethel Ong	Associate Professor, De La Salle University, Philippines

11:30 – 11:40 Mr Shehzad Mevawalla Director Alexa Speech Recognition, Amazon, USA
Alexa, how do language technologies work?

11:40 – 11:50 Mr Kelly Davis Machine Learning Group Manager, Mozilla, Germany
Common Voice and Deep Speech: Democratizing Speech Corpora and Tools

11:50 – 12:00 Mr Dmitriy Genzel Research scientist manager, Facebook, Inc., USA
Large-scale content understanding on Facebook

12:00 – 12:10 Mr Radu Florian Senior Manager and Distinguished Research Staff Member, IBM Research AI, (USA)
The Journey to Multilingual Watson: Best Practices for Building NLP Models for New Languages

12:10 – 12:20 Mr Shijin Wang Dean of iFLYTEK AI Research, IFLYTek, China
How do Speech Translation Technologies work in iFLYTEK Translator

12:20 – 12:30 Mr Amit Kumar Pandey President, Chief Science Officer (CSO), Chief Technology Officer (CTO), Hanson Robotics, China
Language Technologies, Social Robots and the Society-
The needs, challenges and the impact

12:30-13:00 Panel discussion

13:00-14:30 Poster Session 2: European and Arctic Languages
During Lunch Break - Poster Area Hall Segur

Moderators/Rapporteurs:

Ms Teresa Lynn	Research Fellow, ADAPT Centre, Dublin City University, Ireland
Mr Daniil Kocharov	Associate Professor, Saint Petersburg State University, Russian Federation
Mr Vicent Fenollar	Policy and Outreach Manager, Network to Promote Linguistic Diversity (NPLD), Belgium

P.2.1 Mr Peter Bouda Poio

Mr Peter Bouda

Poio - Open Source Technology for Language Diversity

P.2.2 Mr Thierry Declerck DFKI GmbH

Mr Simon Krek, Mr Thierry Declerck, Mr John Philip McCrae and Ms Tanja Wissik

Towards a Global Lexicographic Infrastructure

P.2.3 Mr Laurent Kevers Università di Corsica Pasquale Paoli

Mr Laurent Kevers, Ms Stella Retali-Medori, Mr Florian Guéniot and Ms A. Ghjacumina Tognotti

Tooling up a less-resourced language with NLP : the example of Corsican and the "Banque de Données Langue Corse" (BDLC, Corsican Language Database)

P.2.4 Ms Anna Nikulásdóttir Grammatek ehf

Ms Anna Nikulásdóttir

Language Technology Program for Icelandic

P.2.5 Mr Philippe Boula de Mareüil LIMSI-CNRS

Mr Philippe Boula de Mareüil, Mr Gilles Adda, Mr Albert Rilliard and Mr Frédéric Vernier

A speaking atlas of indigenous languages of France and its Overseas

P.2.6 Mr Aidar Khusainov Institute of Applied Semiotics of the Tatarstan Academy of Sciences

Mr Dzhavdet Suleymanov, Mr Aidar Khusainov and Mr Rinat Gilmullin

Software and Linguistic Resources for the Tatar language preservation and development: Regional Experience

P.2.7 Ms Victoria Bobicev Technical University of Moldova

Ms Victoria Bobicev, Ms Catalina Mărănduc, Mr Tudor Bumbu, Ms Ludmila Malahov, Mr Alexandru Colesnicov and Ms Svetlana Cojocaru

Contribution to the Universal Dependencies Treebank of Non-Standard Romanian Texts

P.2.8 Ms Claudia Soria Consiglio Nazionale delle Ricerche - Istituto di Linguistica Computazionale "A. Zampolli"

Ms Claudia Soria and Mr Cor van der Meer

Inquiring about digital use and usability of minority languages: the approach of the Digital Language Diversity Project

P.2.9 Ms Patricia Serbac University of Tîrgu-Mureș

Ms Patricia Serbac

Language Technologies for Istro-Romanian

P.2.10 Ms Birna Arnbjörnsdóttir Vigdis Institute, University of Iceland

Ms Birna Arnbjörnsdóttir and Ms Auður Hauksdóttir

Innovative CALL Solutions and the Sustainability of Nano Languages in the North West Arctic Region

P.2.11 Mr Brendan Molloy The Techno Creatives

Mr Brendan Molloy

Indigenous/Minority Language Keyboard and Spell Checking Support, for Desktop and Mobile Operating Systems

P.2.12 Mr Damien Nouvel Inalco ERTIM

Mr Damien Nouvel, Mr Driss Sadoun and Mr Mathieu Valette

MultiTAL : an online platform to list NLP tools for under-resourced languages

P.2.13 Mr Valeriy Pylypenko Speech Science and Technology Department International Research/Training Center for Information Technologies and Systems, Kyiv

Mr Valeriy Pylypenko and Ms Tetyana Lyudovyk

Automatic Recognition of mixed Ukrainian-Russian Speech

P.2.14 Mr Mikel L. Forcada Universitat d'Alacant

Mr Mikel L. Forcada and Mr Francis Tyers

Apertium: a free/open-source platform for machine translation and basic language technology

P.2.15 Ms Oana Niculescu The Romanian Academy Institute of Linguistics “Iorgu Iordan – Al. Rosetti”

Ms Oana Niculescu, Ms Maria Marin and Ms Daniela Răuțu

Rediscovering past narrations: the oral history of the Romanian language preserved within the national phonogramic archive

P.2.16 Mr Trond Trosterud U Tromso

Mr Sjur Moshagen, Ms Lene Antonsen and Mr Trond Trosterud

Language technology for indigenous languages: Achievements and challenges

P.2.17 Ms Tracey Herbert First Peoples Cultural Council

Ms Lorna Williams, Ms Tracey Herbert and Mr Daniel Yona

Using technology to empower Indigenous knowledge sharing

P.2.18 Ms Sabine Kirchmeier Dansk Sprognævn/Danish Language Council & Kirchmeier.dk

Ms Sabine Kirchmeier

European Language Monitor by EFNIL

P.2.19 Ms Amel Fraisse Université de Lille

Ms Amel Fraisse, Mr Ronald Jenn, Ms Shelley Fisher Fishkin and Mr Zheng Zhang

Preserving Endangered European Cultural Heritage and Languages Through Translated Literary Texts

P.2.20 Ms Benedicte Haraldstad Frostad The Language Council of Norway

Ms Benedicte Haraldstad Frostad

Towards ASR that recognises everyone in a country with no spoken standard

P.2.21 Mr Jack Rueter University of Helsinki, Digital Humanities

Mr Jack Rueter

Komi Latin-Alphabet Letters Not Found in Unicode

P.2.22 Mr Niko Partanen University of Helsinki

Mr Niko Partanen, Mr Michael Rießler and Mr Thierry Poibeau

Developing technologies for low-resource Uralic languages: Case studies on Saami and Komi varieties

P.2.23 Ms Maria Eskevich CLARIN ERIC

Ms Maria Eskevich and Ms Franciska de Jong

Understanding culture and society with the language resources and tools offered through the CLARIN Research Infrastructure

P.2.24 Mr Alexey Karpov St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences

Mr Alexey Karpov, Mr Ildar Kagirov, Mr Dmitry Ryumin and Mr Alexander Axyonov

A Multimodal Database of Russian Sign Language

P.2.25 Mr Mikkel Rasmus Logje The Sámi Parliament

Mr Mikkel Rasmus Logje

Sámi languages

P.2.26 Mr Tamás Váradi MTA Research institute for Linguistics

Mr Marko Tadić and Mr Tamás Váradi

LT Data Free for All

P.2.27 Ms Jacqueline Brixey USC Institute for Creative Technologies

Ms Jacqueline Brixey, Mr Seyed Hossein Alavi and Mr David Traum

Can we use a spoken Dialogue System to document Endangered Languages?

P.2.28 Ms Adrià Martín-Mor Universitat Autònoma de Barcelona

Ms Adrià Martín-Mor

Technologies for Endangered Languages: The Case of the Languages of Sardinia

14:30-16:15 Session 6: Scientific aspects related to the state of the art in Language Technologies, for spoken, written and sign languages

Moderators:

Mr Volker Steinbiss	General Manager AppTek, Germany
Ms Ximena Guttierez-Vasques	Postdoctoral Researcher, Universität Zürich, Switzerland

***Rapporteur*:**

Ms Dessi Puji Lestari	Head of Research and Development of Artificial Intelligence Research Centre, Bandung Institute of Technology / Prosa.ai, Indonesia

14:30-14:40	Ms Judith Klavans	Senior Research Scientist, University of Maryland, USA

Achievements and Contributions from Natural Language Processing (NLP)

14:40-14:50	Mr Jean Senellart	CEO, Systran, France

NMT4All (Neural Machine Translation for All)

14:50-15:00	Mr Alex Waibel	Professor, Director, Karlsruhe Institute of Technology, Institute for Anthropomatics and Robotics, Germany
	Mr Sebastian Stüker	Associate Fellow, Karlsruhe Institute of Technology (KIT), and CEO, Karlsruhe Information Technology Solutions – kites GmbH

Communicating with Everybody: Going the Extra Mile toward Automatic Simultaneous Interpretation for All

15:00-15:10	Ms Aijun Li	Professor, Institute of Linguistics, Chinese Academy of Social Sciences, China

Handling Prosody and Tone Languages

15:10-15:20	Mr Ahmed Ali	Principal Engineer, Qatar Computing Research Institute, Hamad Ben Khalifa University, Qatar

Dialectal Speech Processing: Past, Present, Future

15:20-15:30	Mr Roger Moore	Professor of Spoken Language Processing, University of Sheffield, UK

Talking with Robots: Opportunities and Challenges

15:30-15:40	Mr Thomas Hanke	Researcher, Universität Hamburg, Germany

Recent Advances on Sign Language Technologies

15:40-16:15 Panel Discussions

16:15-16:45 Poster Session 3: Latin America and the Caribbean Languages
Coffee Break - Poster Area Hall Segur

***Moderators/Rapporteurs*:**

Mr Francisco Cláudio Samapaio de Menezes	Adjunct Professor, University of Brasilia, Brazil
Mr Marco Antonio Martínez Pérez	Activist, Kumoontun, Mexico
Ms Anuschka van ‘t Hooft	Professor, Autonomous University of San Luis Potosí, Mexico

P.3.1 Mr Jhonnatan Rangel INALCO-SeDyL

Jhonnatan Rangel

Challenges for language technologies in Ayapaneco

P.3.2 Mr Michael Gasser Indiana University

Mr Michael Gasser

Mainumby: computer-assisted Spanish-to-Guarani translation

P.3.3 Mr Maximiliano Duran Université Franche-Comté

Mr Maximiliano Duran

Baby Quechua robot

P.3.4 Ms Cynthia Montaño Universidad Nacional Autónoma de México

Ms Cynthia Montaño, Mr Gerardo Sierra Martínez and Ms Gemma Bel-Enguix

On the development of the Mexican Languages Parallel Corpus

P.3.5 Ms Tajëëw Díaz Colmix

Ms Tajëëw Díaz

Project: Endless Oaxaca Multilingual

P.3.6 Ms Alejandrina Cristia Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL Research University

Ms Camila Scaff, Mr Marvin Lavechin and Ms Alejandrina Cristia

Large-scale audio-recordings to study infant language acquisition

P.3.7 Ms Vania Ramírez UNAM

Ms Vania Ramírez

Nierika Red Social para aprender y enseñar una lengua indígena

P.3.8 Ms Purvi Shah Pratham Books

Ms Purvi Shah

Preserving indigenous languages in south and central America by leveraging open licensing and technology

P.3.9 Ms Ximena Gutierrez-Vasques UNAM

Ms Ximena Gutierrez-Vasques and Mr Victor Mijangos

Comunidad Elotl. Language Technologies for Mexico's Indigenous Languages

P.3.10	Ms Emiliana Cruz	CIESAS-CDMX

Ms Emiliana Cruz
Language and Landscape: Hiking and Documenting the Chatino Language of San Juan Quiahije

P.3.11	Mr Luis Flores Martínez	National Institute of indigenous languages of Mexico (INALI)

Mr Luis Flores Martínez
Resources and digital materials in Mexico's indigenous languages

P.3.12	Mr Marco Martinez	Kumoontun

Mr Marco Martinez
Ayöök, México

16:45-18:15 Session 7: Infrastructural aspects

Moderators:

Mr Mark Liberman	Professor, University of Pennsylvania, USA
Ms Amanda Harris	Research Fellow, Director of PARADISEC Sydney unit PARADISEC, University of Sydney, Australia

Rapporteur:

Mr Alexey Karpov	Head of the Speech & Multimodal Interfaces Lab. St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, Russian Federation

16:45-16:55	Ms Denise di Persio	Associate Director, Linguistic Data Consortium University of Pennsylvania, USA

The Linguistic Data Consortium: Developing and Distributing Language Resources4All

16:55-17:05	Mr Yohei Murakami	Associate Professor, Risumeikan University, Japan

Language Sphere: A Socio-Technical Approach to Bilingual Dictionary Creation for Indigenous Languages

17:05-17:15	Mr Craig Cornelius	Senior Software Engineer, International Engineering Google, Inc. USA

Every Language on the Web - How Unicode makes it happen!

17:15-17:25	Mr Carl Rubino	Research Program Manager, IARPA, USA

IARPA's Contribution to Low Resource HLT Development and Evaluation

17:25-17:35	Mr Sanjeev Khudanpur	Associate Professor, The Johns Hopkins University, USA

Open source toolkits and Tutorial/Summer school

17:35-17:45	Mr Georg Rehm	Principal Researcher and Research Fellow, DFKI GmbH, Germany

European Language Grid: Language Technologies for Europe

17:45-18:15 Panel discussion

18:30-20:30 Social event (Welcome Reception Hall Segur)

Moderators/Rapporteurs:

Ms Sakriani Sakti	Associate Professor, Nara Institute of Science and Technology, Japan Secretary of SIGUL, ELRA-ISCA Special Interest Group of Under-Resourced Languages, France
Mr Joseph Mariani	Senior Researcher, LIMSI-CNRS, France; Honorary President of ELRA, France
Mr Khalid Choukri	Secretary General, European Language Resources Association / ELDA CEO, France

Special event (1)	Mr Tunde Adegbola	Executive Director, African Languages Technology Initiative (Alt-i), Nigeria

Yoruba Drum Language: Implications for Language Technology

Special event (2)	Mr Rory O'Connor	Chief Executive Officer, Yugambeh Museum, Australia

How Yugambeh's Aboriginal language revitalisation has helped shape Woolaroo, an App for all Indigenous languages

Programme 6 December 2019

Day 3. CHALLENGES: Addressing the digital divide and multilingualism

(Room II and Poster area: Hall Segur)

09:00-09:30 Session Keynote 2:

Moderators:

Ms Tanja Shultz	Professor, University of Bremen, Cognitive Systems Lab, Germany
Mr Justus Roux	Researcher, South African Centre for Digital Language Resources (SADiLaR), South Africa

Rapporteur:

Mr Priyankoo Sarmah	Associate Professor, Indian Institute of Technology Guwahati, India

09:00-09:30	Ms Lorna Williams	Board Member; Prof Emerita, First Nations Cultural Foundation; University of Victoria, Canada

Wa7 szum'in'stum' ti nqwelutenlhkalha - Technology and Indigenous Language Revitalization, Recovery and Normalization

---

09:30-11:00 Session 8: Minority and Indigenous Languages

Moderators:

Mr Yoshinori Sagisaka	Professor, Waseda University, Japan
Ms Sonja Bosch	Professor, University of South Africa, South Africa

Rapporteur:

Ms Claudia Soria	Researcher, Consiglio Nazionale delle Ricerche, Italy

09:30-09:40	Mr Borja L.C. Patrocino Antonino	Universidade Nacional de Timor Leste, Timor Leste

Speech Technology for Indigenous Language in Timor Leste

09:40-09:50	Ms Evelyn Fogwe Chibaka	Professor of Linguistics, University of Buea, Cameroon

Language Technology Applications in Africa within an "Inclusive, innovative and reflective" crisis Interface

09:50-10:00	Ms Fadoua Ataa Allah	Director, Centre des Etudes Informatiques, IRCAM, Morocco

The IRCAM Realizations for the Amazigh Preservation and Revitalization in Morocco

10:00-10:10 Ms Aili Keskitalo President of the Sami Parliament and Co-chairs of the Steering Committee of International Year of Indigenous Languages, Norway

The voices of Indigenous Languages in the infinite cyberspace

10:10-10:20 Ms Cecilia Piaggio Founder / Program Director, Latinoamérica Habla / RWS Moravia, Argentina

Our learnings when rolling out tech initiatives in Qom

10:20-10:30 Ms Apolonia Tamata Senior Culture & Heritage Specialist, iTaukei Trust Fund, Fiji

Challenges with indigenous minority languages and language technologies

---

10:30-11:00 Panel discussion

11:00-11:30 Poster Session 4: African Languages
Coffee Break - Poster Area Hall Segur

***Moderators/Rapporteurs**:*

Ms Dorothy Beermann	Professor, Norwegian University of Science and Technology, Norway
Ms Evelyn Fogwe Chibaka	Professor of Linguistics, University of Buea, Cameroon
Mr Audace Niyonkuru	Chief Executive Officer, DIGITAL UMUGANDA, Rwanda

P.4.1 Ms Marissa Griesel University of South Africa

Ms Sonja Bosch and Ms Marissa Griesel

African Wordnet – digital documentation and preservation of indigenous knowledge

P.4.2 Ms Brigitte BIGI LPL, CNRS, Aix-en-Provence, (France)

Ms Brigitte BIGI

Automated Speech Segmentation: Example of an African Language

P.4.3 Z Steyn SADiLaR

Z Steyn

Establishing Sustainable Infrastructures for African Languages

P.4.4 Ms Febe De Wet Stellenbosch University

Ms Febe De Wet, Mr Ewald Van der Westhuizen and Mr Thomas Niesler

A South African Corpus of Multilingual Code-switched Soap Opera Speech

P.4.5 Mr Valentin Vydrin INALCO

Mr Valentin Vydrin

Corpora Mandeica: text corpora for Mande languages (West Africa)

P.4.6 Ms Kerry Jones African Tongue and Stellenbosch University

Ms Kerry Jones

Missing link: A centralised digital archive for endangered languages of southern Africa

P.4.7 Mr Christopher Cieri Linguistic Data Consortium, University of Pennsylvania
Mr Christopher Cieri and Mr Mark Liberman
Using Citizen Linguistics to Empower Indigenous Communities

P.4.8 Mr Lulamile Mzamo North-West University
Mr Lulamile Mzamo, Mr Albert Helberg and Ms Sonja Bosch
Heuristic guided probabilistic graphic language modelling for morphological segmentation of isiXhosa

P.4.9 Ms Febe De Wet Stellenbosch University
Mr Astik Biswas, Ms Febe De Wet, Mr Herman Kamper, Raghav Menon, Mr Thomas Niesler, Mr Armin Saeb, Mr John Quinn, Mr Ewald Van der Westhuizen and Emre Yilmaz
Radio-browsing in support of relief and development work in rural Africa

P.4.10 Ms Martha Yifiru Tachbelie Addis Ababa University
Ms Martha Yifiru Tachbelie, Mr Solomon Teferra Abate and Ms Tanja Schultz
Analysis of Language Relatedness for the Development of Multilingual Automatic Speech Recognition for Ethiopian Languages

P.4.11 Mr Lucas Ondel Brno University of Technology
Mr Lucas Ondel and Mr Lukas Burget
Automatic Learning of a Phonological System: a Case Study on the Mboshi Language

P.4.12 Mr Seid Yimam Universität Hamburg
Mr Seid Yimam and Mr Chris Biemann
Current Status, Issues, and Future Directions for Ethiopian Natural Language Processing (NLP) Research

P.4.13 Mr Martin Benjamin Kamusi Project International
Mr Martin Benjamin
ACALAN: Platform for African Language Empowerment (PALE)

P.4.14 Mr Moses Ekpenyong University of Uyo
Mr Moses Ekpenyong, Ms Eno-Abasi Urua and Mr Aniefon Akpan
SCAnnAL – An Automatic Speech Corpus Annotator for African Speech Corpora

P.4.15 Mr Damien Nouvel Inalco ERTIM
Mr Elvis Mboning and Mr Damien Nouvel
NTeALan - Artificial Intelligence, Development and Promotion of African National Languages

P.4.16 Mr Sunday Ojo Tshwane University of Technology
Mr Sunday Ojo
Grappling with Opportunities and Challenges in developing Language Technologies for Under-resourced African Languages

P.4.17 Ms Fatiha Sadat UQAM
Mr Houssein Ahmed Assowe and Ms Fatiha Sadat
Towards the First Low-Resource French-Somali Machine Translation System

P.4.18 Ms Fatiha Sadat UQAM
Bilal Belainine and Ms Fatiha Sadat
Automatic Dialect Detection in Arabic Broadcast Media

P.4.19 Mr Abdoulaye Barry Winden Jangen
Mr Abdoulaye Barry and Mr Ibrahima Barry
The ADLaM Story

11:30-13:00 Session 9: Activities for language preservation, reclamation, and enhancement

Moderators:

Ms Cyntia Montaño	Project Manager, Grupo de Ingeniería Lingüística, Mexico
Mr Mikel Lorenzo Forcada Zubizarreta	Professor, Universitat d'Alacant, Spain

Rapporteur:

Mr Satoshi Tamura	Associate Professor, Gifu University, Japan

11:30-11:40 Mr Steven Bird Professor, Charles Darwin University, Australia
Designing for Language Revitalisation

11:40-11:50 Ms Mandana Seyfeddinipur Endangered Languages Documentation Programme, Director, SOAS University of London, UK
We just scratched the surface: 16 years of supporting endangered languages documentation

11:50-12:00 Mr Alan Black Professor, Carnegie Mellon University, USA
Using found data to build speech recognition and speech synthesis for all languages

12:00-12:10 Mr Antti Arppe Associate Professor, University of Alberta, Canada
21st century language technology and tools meet 21st century challenges and opportunities

12:10-12:20 Ms Jeannette Stewart Founder, Translation Commons, USA
Translation Commons: No Language and No Linguist Left Behind

12:20-12:30 Mr David Traum Research Professor, University of Southern California, USA
Conversations with History: Spoken dialogue technology to preserve culture and language

12:30-13:00 Panel discussion

13:00-14:30 Poster Session 5: Asian Languages
Lunch Break - Poster Area Hall Segur

Moderators/Rapporteurs:

Ms Budi Irmawati	Lecturer, University Mataram, Indonesia
Mr Lorang Yun	Secretariat Coordinator, Cambodia Indigenous Peoples Alliance, Cambodia
Mr Netra Mani Rai	Language Development Project Adviser, International Nepal, Nepal

P.5.1 Mr Thierry Declerck — DFKI GmbH

Mr John McCrae and Mr Thierry Declerck

Linguistic Linked Open Data for All

P.5.2 Mr Mohammad Nurul Huda — United International University and eGeneration Ltd.

Mr Mohammad Nurul Huda

Bangla Text and Spoken Language Technology

P.5.3 Mr Bal Krishna Bal — Department of Computer Science and Engineering, Kathmandu University, Nepal

Mr Bal Krishna Bal, Mr Amrit Yonjan Tamang and Mr Lasang Jimba Tamang

Envisioning a Trilingual Machine Translation System for the Language Pairs – <Tamang –English –Nepali>

P.5.4 Mr Marc Durdin — SIL International

Mr Marc Durdin, Mr Sok Makara, Mr Joshua Horton and Ty Rasmey

Keyman: High Fidelity Text Input for All Languages

P.5.5 Ms Natsuko Nakagawa — National Institute for Japanese Language and Linguistics

Ms Natsuko Nakagawa, Mr Masahiro Yamada, Mr Kenan Celik, Ms Nobuko Kibe and Mr Yukinori Takubo

Digital archiving and museum for language documentation and revitalization in Japan

P.5.6 Ms Sunayana Sitaram — Microsoft Research India

Ms Sunayana Sitaram, Mr Monojit Choudhury and Ms Kalika Bali

Project Mélange: Speech and Language Technologies for Code-switching

P.5.7 Mr Martin Raymond — SIL International

Mr Martin Raymond and Mr Peter Martin

Providing smart, open fonts for the world's language communities

P.5.8 Ms Dessi Puji Lestari — Institut Teknologi Bandung

Ms Ayu Purwarianti, Ms Dessi Puji Lestari and Mr Teguh Eko Budiarto

InaNLP: Indonesian Natural Language Processing Tools API

P.5.9 Mr Alexis MICHAUD CNRS - LACITO
Ms Séverine Guillaume, Mr Balthazar Do Nascimento and Mr Alexis MICHAUD
The Pangloss Collection: an open archive of under-documented languages designed with Natural Language Processing in view

P.5.10 Mr Virach Sornlertlamvanich Faculty of Data Science, Musashino University
Mr Virach Sornlertlamvanich, Mr Teguh Eko Budiarto and Ms Thatsanee Charoenporn
Multi-lingual Support in Connective Learning Scheme for Refining and Connecting the Open Educational Videos

P.5.11 Ms Ethel Ong De La Salle University
Ms Ethel Ong, Ms Nathalie Rose Lim-Cheng, Ms Charibeth Cheng and Mr Edward Tighe
Promoting and Preserving Philippine Culture and Languages through Language Technologies

P.5.12 Mr Virach Sornlertlamvanich Faculty of Data Science, Musashino University
Ms Thatsanee Charoenporn and Mr Virach Sornlertlamvanich
Improvement of Thai NER and the Corpus

P.5.13 Ms Kalika Bali Microsoft Research Labs
Ms Kalika Bali, Mr Monojit Choudhury, Ms Sunayana Sitaram and Mr Sebastin Santy
Deploying Language Technologies for Underserved Communities

P.5.14 Ms Angelina Aquino University of the Philippines Diliman
Ms Angelina Aquino and Mr Rhandley Cajote
Language Technologies at the University of the Philippines DIliman

P.5.15 Mr Tenzin Namgyel Dzongkha Development Commission
Mr Tenzin Namgyel
Languages and Technology in Bhutan

P.5.16 Mr Ramakrishnan Angarai Ganesan Indian Institute of Science
Mr Ramakrishnan AngaraiGanesan
Language technology at MILE Lab, Indian Institute of Science

P.5.17 Mr Udaya Narayana Singh Amity University Haryana
Udaya Narayana Singh
Multilingual profile of India

P.5.18 Mr Udaya Narayana Singh Amity University Haryana
Mr Udaya Narayana Singh, Ms Esha Jainiti, Ms Rusha Mudgal and Ms Anwita Maiti
Mediating Multilingualism

P.5.19 Mr Vijay Kumar Ministry of Electronics and Information Technology
Vijay Kumar and Dr S K Srivastava
Technology Development for Indian Languages

P.5.20 Ms Purvi Shah Pratham Books
Ms Purvi Shah

Creating access to openly licensed early reading resources in Asia's indigenous languages

P.5.21 Mr Virach Sornlertlamvanich — Faculty of Data Science, Musashino University

Ms Ari Yanase, Ms Thatsanee Charoenporn and Mr Virach Sornlertlamvanich

Conversational Bot for Eyesight Testing Automation

P.5.22 Mr Arbi Haza Nasution — Universitas Islam Riau

Mr Arbi Haza Nasution and Mr Totok Suhardijanto

Dictionary 4.0: Alternative Presentations for Indonesian Multilingual Dictionaries

P.5.23 Ms Viyazonuo Terhiija — Indian Institute of Technology Guwahati

Ms Viyazonuo Terhiija, Mr Samudra Vijaya and Mr Priyankoo Sarmah

Speech Technology in three tonal languages of North-East India

P.5.24 Ms Win Pa Pa — University of Computer Studies, Yangon

Ms Win Pa Pa

Bringing Zero-resourced Languages of Myanmar to the Digital World

P.5.25 Mr Totok Suhardijanto — University of Indonesia

Mr Totok Suhardijanto and Ms Arawinda Dinakaramani

Building Corpora for Under-Resourced Languages in Indonesia

P.5.26 Mr Ritesh Kumar — Department of Linguistics, Dr. Bhimrao Ambedkar University, Agra

Mr Ritesh Kumar, Ms Bornini Lahiri, Mr Atul Kr. Ojha, Mr Mayank Jain and Mr Deepak Alok

Language Resources and Technology Development Efforts for some Lesser-known Indian Languages

P.5.27 Mr David Yarowsky — Johns Hopkins University

Mr David Yarowsky, Mr Arya D. McCarthy, Mr Garrett Nicolai, Mr Winston Wu, Mr Aaron Mueller, Mr Dylan Lewis, Ms Yingqi Ding, Mr Abhinav Nigam, Mr Emre Ozgu, Mr Debanik Purkayastha, Mr James Scharf and Mr Kenneth Zheng

A 1000-language Collaborative Universal Dictionary and Universal Translator

P.5.28 Mr Nathaniel Oco — De La Salle University

Mr Nathaniel Oco

Tagalog-English Code-Switching: Challenges for Automatic Detection

P.5.29 Ms Fatiha Sadat — UQAM

Mr Tan Ngoc Le and Ms Fatiha Sadat

How a low-resource named entities recognition and transliteration framework for Vietnamese can improve the automatic machine translation ?

P.5.30 Mr Shyam Sundar Agrawal — KIIT College of Engineering, Gurgaon

Ms Shweta Sinha and Mr Shyam Sundar Agrawal

Situation and challenges of technologies for indigenous languages of India

P.5.31 Mr Craig Cornelius — Google, Inc.

Mr Craig Cornelius

Unicode for Indigenous Languages - Standards and technology for getting online

P.5.32	Wei Wang	Institute of Linguistics, Chinese Academy of Social Sciences

Wei Wang, Ms Aijun Li and Ms Danqing LIU

CASS-LING's Linguistic Infrastructure: Resources, Platforms and Services

14:30-16:00 Session 10: Scientific aspects related to handling language diversity

Moderators:

Mr Laurent Besacier	Professor, LIG - University Grenoble Alpes (France)
Mr Ossama Emam	IBM Senior Technical Staff Member, IBM Egypt, (Egypt)

Rapporteur:

Mr Nathaniel Oco	De La Salle University (Philippines)

14:30-14:40	Ms Odette Scharenborg	Associate Professor, Delft University of Technology, Netherland

Towards speech technology for unwritten languages

14:40-14:50	Mr Satoshi Nakamura	Professor, Nara Institute of Science and Technology, Japan

Semi-supervised learning: Machine Speech Chain, Code-switch & Automatic Interpretation

14:50-15:00	Mr TrondTrosterud, Mr Sjur Nørstebø Moshagen	Professor of Saami Language Technology Chief Engineer, UiT The Arctic University of Norway, Norway

Rich morphology, no corpus - and we still made it. The Sámi experience

15:00-15:10	Mr Holger Schwenk	Professor, Facebook, France

Mining for Multilingual Resources on the WEB

15:10-15:20	Mr Brian Roark	Research Scientist, Google, Inc. USA

Text entry for All

15:20-15:30	Ms Kalika Bali	Principal Researcher, Microsoft Research Labs India, India

Evaluating Speech Data and Technology for Indian Languages

15:30-16:00 Panel discussion

16:00-16:30 Poster Session 6: North America Languages
Coffee Break - Poster Area Hall Segur

***Moderators/Rapporteurs**:*

Mr Chris Cieri	Adjunct Associate Professor, Linguistics; Executive Director, Linguistic Data Consortium, University of Pennsylvania, USA
Mr Francis Tyers	Assistant Professor, Indiana University, USA
Mr Jan Trmal	Associate Research Scientist, Johns Hopkins University, USA

P.6.1 Ms Lynne Bowker — University of Ottawa

Ms Lynne Bowker

Machine Translation 4 All: Developing informed and critical users through a program of machine translation literacy

P.6.2 Ms Marie-Odile Junker — Carleton University

Ms Marie-Odile Junker and Delasie Torkornoo

Building a common Digital Infrastructure to sustain Algonquian Languages

P.6.3 Ms Emily Prud'hommeaux — Boston College

Ms Emily Prud'hommeaux, Mr Robert Jimerson, Mr Richard Hatcher, Mr Raymond Ptucha and Ms Karin Michelson

On the promise and pitfalls of repurposing existing language technologies for endangered language documentation

P.6.4 Ms Alexa Little — 7000 Languages

Ms Alexa Little, Ms Kayleigh Jeannette and Ms Kelsey Riggs

7000 Languages: Free Language-Learning Software for Language Reclamation

P.6.5 Mr Aidan Pine — National Research Council Canada

Mr Aidan Pine, Mr Nathan Brinklow, Ms Heather Souter and Ms Delaney Lothian

National Research Council Canada Indigenous Language Technology Project

P.6.6 Ms Lane Schwartz — University of Illinois at Urbana-Champaign

Ms Lane Schwartz, Ms Emily Chen, Ms Hayley Park, Ms Sylvia Schreiner and Mr Benjamin Hunt

St. Lawrence Island Language Technology for Documentation & Revitalization

P.6.7 Delaney Lothian — University of Alberta

Ms Delaney Lothian, Ms Daniela Teodorescu, Mr Denilson Barbosa and Ms Carrie Demmans Epp

Building a Language Model of Nehiyawewin (Cree, Y-dialect)

P.6.8 Mr Roy Boney — Cherokee Nation

Mr Roy Boney
From Talking Leaves to Pixels: The Evolution of the Cherokee Syllabary

P.6.9 Ms Jacqueline Brixey USC Institute for Creative Technologies

Ms Jacqueline Brixey
ChoCo: A multimodal corpus for the Choctaw language

P.6.10 Ms Fatiha Sadat UQAM

Ms Fatiha Sadat, Mr Tan Ngoc Le and Mr David Huggins Daines
Issues and challenges of NLP in relation to Canada's Aboriginal languages

---

16:30-17:30 Session 11: Developing Language Technologies for All: Best Practices

Moderators:

Ms Delyth Prys	Head of Language Technologies, Bangor University, UK
Ms Astrid Berengere Mengue	Project Coordinator, Center for Human Rights and Democracy in Africa, Cameroon

Rapporteur:

Ms Emiliana Cruz	Researcher, CIESAS-CDMX, Mexico

16:30-16:40 S.K. Srivastava Senior Director/Scientist G, Government of India, Ministry of Electronics and IT, India
Technology Development for Indian Languages: Government Initiatives

16:40-16:50 Mr Juan Steyn Project Manager
South African Centre for Digital Language Resources
South Africa
The South African Research Infrastructure Roadmap and national mandate of the South African Centre for Digital Language Resources

16:50-17:00 Mr Vaine Tutai Richard Language Programme Lead, Ministry for Pacific Peoples, New Zealand
Lalanga Fou: weaving a new collaborative and regional approach to ensure the survival of Pacific languages, cultures and identities in Aotearoa New Zealand and the Pacific Oceania region through technology

17:00-17:10 Mr Aodhán Mac Cormaic Director, Ministry for Culture, Heritage and the Gaeltacht, Ireland
Language technologies in the language planning framework for Irish

17:10-17:20 Mr Norberto Zamora Chief of Department, Instituto Nacional de los Pueblos Indígenas - Departamento de Medios Digitales, Mexico
Códice México

17:20-17:30 Ms Tracey Herbert Chief Executive Officer, First Peoples' Cultural Council, Canada
Indigenous perspective on developing and maintaining an Indigenous curated Technology to support language revitalization

17:30-18:30 Session 12: The Future of Language Technologies for All: Outcome Document, Recommendations, and Closing remarks

Moderators and Rapporteurs **:**

Mr Moez Chakchouk	Assistant Director-General for Communication and Information, UNESCO
Mr Khalid Choukri	Secretary General, European Language Resources Association / ELDA CEO
Mr Gilles Adda	Researcher, LIMSI-CNRS, France Member of the ELRA Board
Rapporteurs from Day One	

Proceedings of the Language Technologies for All (LT4All) , pages 147–150
Paris, UNESCO Headquarters, 5-6 December, 2019.

Dictionary 4.0: Alternative Presentations for Indonesian Multilingual Dictionaries

Arbi Haza Nasution, Totok Suhardijanto
Informatics Engineering Department Universitas Islam Riau, Linguistics Department Universitas Indonesia
Pekanbaru Riau Indonesia, Jakarta Indonesia
arbi@eng.uir.ac.id, totok.suhardijanto@ui.ac.id

Abstract

Building a multilingual dictionary for 719 languages in Indonesia is a challenging task. We have developed application to create the Leipzig-Jakarta list database for all indigenous languages in Indonesia. The database can be used to generate lexical similarity or lexical distance matrix between languages by comparing the word list. For starter, we covered 11 languages: Indonesian, Javanese, Sundanese, Madurese, Bima, Ternate, Tidore, Palembang Malay, Mandailing Batak, Malay, and Minangkabau. The application has two main features: exploring the existing translations and adding translations to a new language or editing existing translations through crowdsourcing. User acceptance test showed 3.48/4 score.

Keywords: multilingualism, multilingual dictionary, lexical network, lexical computation, computational linguistics

Abstrak

Membangun kamus multibahasa untuk 719 bahasa di Indonesia adalah tugas yang berat. Kami telah mengembangkan aplikasi untuk membuat pangkalan data daftar Leipzig-Jakarta untuk semua bahasa daerah di Indonesia. Pangkalan data tersebut dapat digunakan untuk menghasilkan kesamaan leksikal atau matriks jarak leksikal antar bahasa dengan membandingkan daftar kata tersebut. Sebagai permulaan, aplikasi ini mencakup 11 bahasa: Indonesia, Jawa, Sunda, Madura, Bima, Ternate, Tidore, Melayu Palembang, Batak Mandailing, Melayu, dan Minangkabau. Aplikasi ini memiliki dua fitur utama: menjelajahi terjemahan yang ada dan menambahkan terjemahan ke bahasa baru atau mengedit terjemahan yang ada melalui mekanisme urun daya. Uji keberterimaan pengguna menunjukkan skor 3,48 / 4.

1. Introduction

According to (Eberhard et al., 2019), there are 719 languages in Indonesia, where 707 languages are still alive and 12 languages have become extinct. Extinct in this sense is that there are no longer any of the speakers. Among the surviving languages, 701 languages are local languages and 6 languages are not local languages. Furthermore, there are 18 languages that are used as administrative and / or educational languages, 73 languages are still growing, 188 languages are classified as strong, 347 languages are in difficulty, and 81 languages are in a danger of extinction.

Furthermore, based on his observations, (Anderbeck, 2015) groups Indonesian languages into three groups. First, about two of the four languages in Indonesia today still have a vital life force and have a safe number of speakers (EGIDS (Expanded Graded Intergenerational Disruption Scale) 1-6a). In this group, intergenerational transmission of speakers still occurs and persists. Even though some of them are bilingual, they know when to use local and Indonesian languages. Second, one of the four languages in Indonesia is in fragile condition (EGIDS 6b Threatened) with speakers who continue to decline in number. Usually most young people still learn their mother tongue, but certain reasons make them change their orientation towards languages that are more economically advantageous. Third, the rest, one of the four languages in Indonesia seems to be dying (EGIDS 7-8b) or may have become completely extinct (EGIDS 9 and 10). Some, like the Marori language, may be lost in a generation. The other may be in two or three generations. With conditions like that, of course, we are like racing with time to document language.

Although some experts distinguish the terms of language documentation and language description (Austin and Sallabank, 2011), in some ways, the two are interconnected. According to Austin, the documentation and description of languages differ in their purpose, points of interest, research methods, workflow, and outcomes. Descriptions or language descriptions basically aim at producing grammar, dictionaries, and collections of texts, the target users are generally linguists, and the material produced is sometimes written in a framework that is accessible to trained linguists. In contrast, language documentation is discourse-centered, the main objective being the direct representation of as many types of discourse as possible (Austin, 2007; Woodbury, 2003; Himmelmann, 1998). However, according to (Austin and Grenoble, 2007) the documentation project must rely on the application of theoretical and descriptive linguistic techniques so that the resulting output is sure to be utilized and understood by many communities. So, in other words, documentation and description are activities with objectives and outcomes that complement each other, and one of their important outcomes is the result of lexicographic work, the dictionary.

In the context of endangered languages, dictionaries have a very crucial role, namely storing what is left of endangered languages and cultures by recording valuable information that might be lost (Cristinoi and Nemo, 2013). The bilingual dictionaries are also useful for natural language processing researchers, especially for those related with enrichment of language resources like bilingual dictionary (Nasution et al., 2016; Nasution et al., 2017b; Nasution et al., 2017a; Nasution et al., 2018) or machine translation

(Nasution et al., 2017c; Nasution, 2018). Furthermore, in many cases, the existence of a dictionary can help revive a language and change the attitudes of speakers of that language which ultimately encourage them to use it as often as possible. Even so, (Cristinoi and Nemo, 2013) mentioned that there are some problems related to lexicography in the realm of language documentation. First, the compilers of endangered language dictionaries are generally people or linguists who care. Certainly, the result is different from the general dictionary compiled by a professional team. Secondly, dictionaries made for endangered languages are certainly far from direct economic profit. Third, the endangered language dictionaries have limited distribution, that is only to linguists or the public who have an interest in the language concerned. Fourth, in the work of lexicography in endangered languages there are several problems that must be resolved, for example what characters are used, which variations are considered standard, and so on. Fifth, data collection of endangered languages is more difficult because it only relies on the ethnographic work of researchers or notes from concerned community members. Sixth, the dictionary of threatened languages ??is usually used for research purposes, documenting specific languages and cultures, protecting language and cultural heritage that will be lost without written traditions on the language or culture, helping indigenous people communicate in dominant foreign languages, helping non-native speakers to understand the native speakers and their cultural background, and provide orthography or standard written form for the entire vocabulary.

Because of the problems mentioned above, the data collection of endangered language dictionaries is generally done with a limited number of vocabularies, generally focus on general vocabulary or even basic vocabulary lists. The list is a lexical artifact which is a vocabulary whose references are universally available in many languages in the same region. In the condition of Indonesia which is multilingual, of course the problem becomes more complex. Over time, how do lexicographic studies contribute to language documentation efforts, especially in terms of recording important and varied information about language and culture in Indonesia? Making multilingual dictionaries is not an easy task, especially from the point of computational lexicography (Walker, 1995). Thus, in this paper, we try to build a model that can accommodate the diversity of languages in Indonesia. This can be further elaborated with the question: how to compile dictionaries for languages in Indonesia? What is the correct format of multilingual dictionaries that can help document languages in Indonesia? These two questions will be answered in this paper.

2. Methodology

In the 1950s, linguist Morris Swadesh published a list of 200 words called the Swadesh list, which were thought to be 200 lexical concepts found in all languages ??that were most unlikely to be borrowed from other languages (Swadesh, 1955). Swadesh then reduced the list to 100 items based on intuition where a drastic removal from a 200-word list was the best solution, with the consideration that quality is at least as important as quantity. Although the new list has weaknesses, but the list is relatively light to process because of the small amount. Automated Similarity Judgment Program (ASJP) (Brown et al., 2008) is an open source software with the main objective to develop a Swadesh list database for all languages in the world where lexical similarity or lexical distance matrix between languages can be obtained by comparing the word list. However, the list of 100 Swadesh words was cut down to 40 words that are considered the most stable of forms of change, maintained over time and not replaced by other lexical items from the language itself or elements borrowed from other languages (Holman et al., 2008). The lexical distance between regional languages in Indonesia has been visualized using the ASJP database (Nasution and Murakami, 2019; Nasution et al., 2019). However, there are doubts about the validity of the lexical distance between some regional languages such as between Sundanese and Javanese which should be closer to the lexical distance but only 21.8% of lexical similarities are produced. Therefore, alternative word lists are needed that can produce more accurate lexical distances.

In addition to the Swadesh list, linguists also use the Leipzig-Jakarta list (100 words) (Tadmor et al., 2010) to test the level of chronological separation of languages by comparing words that are resistant to loans. The Leipzig-Jakarta list is available in 2009 (Sakel and Everett, 2012). The mobile application developed in this paper aims to develop the Leipzig-Jakarta list database for all regional languages in Indonesia where lexical similarity or lexical distance matrix between languages can also be further obtained by comparing the word list. The application built will be tested for user satisfaction with quantitative analysis using a questionnaire. The proposed framework is depicted in Figure 1. The data will be used to generate visualization of Indonesian Indigenous Languages Lexical Similarity with Knowledge Graph.

3. Results

For the initial research, 100 Leipzig-Jakarta word lists were translated into 11 languages: Indonesian, Javanese, Sundanese, Madurese, Bima, Ternate, Tidore, Palembang Malay, Mandailing Batak, Malay, and Minangkabau. The application has two main features: exploring the translations of the 100 Leipzig-Jakarta word list and adding translations to new languages or changing translations that are already available. The exploration interface for translating 100 Leipzig word lists into 11 languages with the details of the translated words including the definition, synonyms and example use of the word in a sentence are shown in Figure 2.

To add a translation to a new language or change an already available translation, the user should register to the system first using the registration form. After entering the user's email address for verification, the user can click on the language selection dropdown, then the user can choose the destination language according to the language selection feature, the last step, the user can type the translation according to the language of choice, and click the "*SUNTING / TAMBAH KATA*" (which translated to EDIT / ADD WORDS) button, then the translation added / edited will be

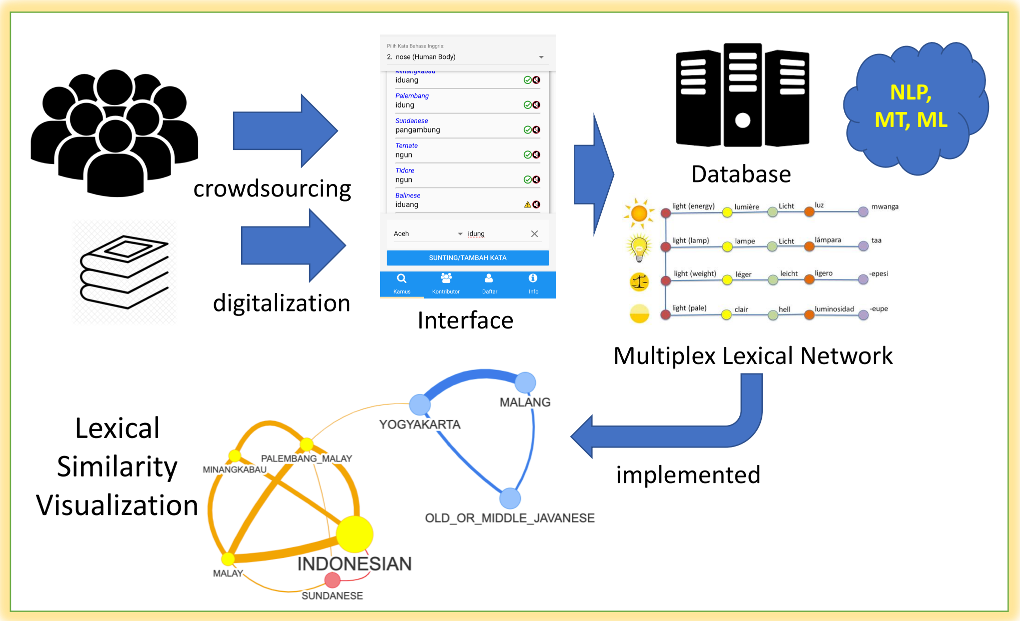
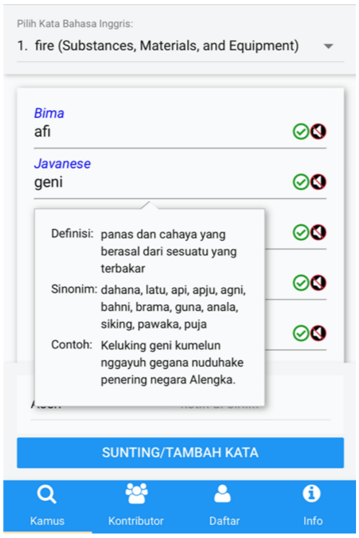


Figure 1: Proposed framework.



Pilih Kata Bahasa Inggris:

1. fire (Substances, Materials, and Equipment)

Bima
afi

Javanese
geni

Definisi: panas dan cahaya yang berasal dari sesuatu yang terbakar

Sinonim: dahana, latu, api, apju, agni, bahni, brama, guna, anala, siking, pawaka, puja

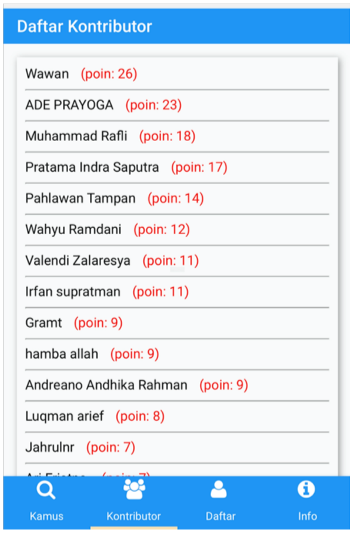
Contoh: Keluking geni kumelun nggayuh gegana nuduhake penering negara Alengka.

SUNTING/TAMBAH KATA

Figure 2: The definition, synonyms and example in sentence.

verified by the linguist. Finally, the user will get a poin for each translation added or edited, and another poin when the new translation or edition has been verified. The leaderboard is shown in Figure 3.

The application that was built was tested by 36 random users with quantitative analysis using a questionnaire with 7 questions as shown in Table 1. Based on the results of the user satisfaction questionnaire with dictionary 4.0, the average value for the whole questionnaire item was 3.48. This shows that the design and appearance of the Dictionary 4.0 Application is quite interesting, easy to use and accepted by users.



Daftar Kontributor

Wawan	(poin: 26)
ADE PRAYOGA	(poin: 23)
Muhammad Rafli	(poin: 18)
Pratama Indra Saputra	(poin: 17)
Pahlawan Tampan	(poin: 14)
Wahyu Ramdani	(poin: 12)
Valendi Zalaresya	(poin: 11)
Irfan supratman	(poin: 11)
Gramt	(poin: 9)
hamba allah	(poin: 9)
Andreano Andhika Rahman	(poin: 9)
Luqman arief	(poin: 8)
Jahrulnr	(poin: 7)

Figure 3: Leader board of contributor.

4. Conclusion

Until now, in this study, a multilingual dictionary prototype model with the functionality to collect data of various languages was quickly compiled. Therefore, the focus in this paper is on the issue of setting up a language data collection system through a crowd sourcing mechanism. Meanwhile, in terms of usage, acceptance testing has been carried out to see how well the application design according to the user. Based on these tests, we obtained quite interesting results, which is 3.48 from a scale of 4. The next stage of this research is to upgrade the dictionary 4.0 application that is capable of managing multilingual dictionary services with dedicated functions for general users and registered users. In addition, the language similarity comparison function

Item	Mean	Median	Standard Deviation
Appealing design and appearance	3.47	3	1.078
The design and appearance of the application is easy to understand	3.41	3	1.043
The navigation menu is easy to understand	3.37	3.5	1.157
The colors used in the application are suitable and not excessive	3.72	4	0.958
The application is easy to use	3.47	4	1.078
Easy to explore each word translation	3.47	4	1.047
It is easy to propose revision to existing translations or add new translations	3.47	3.5	1.047

Table 1: Results of user satisfaction questionnaire of dictionary 4.0

will be included using the lexical distance approach as in the ASJP database program.

5. Acknowledgements

This research was partially supported by Universitas Islam Riau.

6. Bibliographical References

Anderbeck, K. (2015). Portraits of language vitality in the languages of indonesia. *Language documentation and cultural practices in the Austronesian world: Papers from*, pages 19–47.

Austin, P. K. and Grenoble, L. (2007). Current trends in language documentation. *Language documentation and description*, 4:12–25.

Austin, P. K. and Sallabank, J. (2011). *The Cambridge handbook of endangered languages*. Cambridge University Press.

Austin, P. K. (2007). Training for language documentation: Experiences at the school of oriental and african studies. *Documenting and revitalizing Austronesian languages*, pages 25–41.

Brown, C. H., Holman, E. W., Wichmann, S., and Velupillai, V. (2008). Automated classification of the world? s languages: a description of the method and preliminary results. *STUF-Language Typology and Universals Sprachtypologie und Universalienforschung*, 61(4):285–308.

Cristinoi, A. and Nemo, F. (2013). Challenges in endangered language lexicography.

Eberhard, D. M., Simons, G. F., and Fennig, C. D. (2019). Ethnologue: Languages of the world.

Himmelmann, N. P. (1998). Documentary and descriptive linguistics.

Holman, E. W., Wichmann, S., Brown, C. H., Velupillai, V., Müller, A., and Bakker, D. (2008). Explorations in automated language classification. *Folia Linguistica*, 42(3-4):331–354.

Nasution, A. H. and Murakami, Y. (2019). Visualizing language lexical similarity clusters: A case study of indonesian ethnic languages. *Journal of Data Science and Its Applications*, 2(2):45–59.

Nasution, A. H., Murakami, Y., and Ishida, T. (2016). Constraint-based bilingual lexicon induction for closely related languages. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 3291–3298, Paris, France, May.

Nasution, A. H., Murakami, Y., and Ishida, T. (2017a). A generalized constraint approach to bilingual dictionary induction for low-resource language families. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 17(2):9:1–9:29, November.

Nasution, A. H., Murakami, Y., and Ishida, T. (2017b). Plan optimization for creating bilingual dictionaries of low-resource languages. In *2017 International Conference on Culture and Computing (Culture and Computing)*, pages 35–41, Sept.

Nasution, A. H., Syafitri, N., Setiawan, P. R., and Suryani, D. (2017c). Pivot-based hybrid machine translation to support multilingual communication. In *2017 International Conference on Culture and Computing (Culture and Computing)*, pages 147–148, Sept.

Nasution, A. H., Murakami, Y., and Ishida, T. (2018). Designing a collaborative process to create bilingual dictionaries of indonesian ethnic languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 3397–3404, Paris, France, may. European Language Resources Association (ELRA).

Nasution, A. H., Murakami, Y., and Ishida, T. (2019). Generating similarity cluster of indonesian languages with semi-supervised clustering. *International Journal of Electrical and Computer Engineering (IJECE)*, 9(1):1–8.

Nasution, A. H. (2018). Pivot-based hybrid machine translation to support multilingual communication for closely related languages. *World Transactions on Engineering and Technology Education*, 16(2):12–17.

Sakel, J. and Everett, D. L. (2012). *Linguistic fieldwork: A student guide*. Cambridge University Press.

Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *International journal of American linguistics*, 21(2):121–137.

Tadmor, U., Haspelmath, M., and Taylor, B. (2010). Borrowability and the notion of basic vocabulary. *Diachronica*, 27(2):226–246.

Walker, Z. C. (1995). *Automating the lexicon: research and practice in a multilingual environment*. Oxford University Press.

Woodbury, A. C. (2003). Defining documentary linguistics. *Language documentation and description*, 1(1):35–51.