

**Proceedings of the 6th
International Conference on
Natural Language and Speech
Processing (ICNLSP-2023)**

16–17 December, 2023 (virtual)



©2023 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 979-8-89176-065-3

<https://www.icnlsp.org>

Introduction

Welcome to the proceedings of the 6th International Conference on Natural Language and Speech Processing!

This volume presents a vibrant tapestry of cutting-edge research in natural language processing, highlighting advancements in a diverse range of areas. It addresses many NLP aspects as bridging the language divide, expressive and robust communication, building and leveraging resources, and unifying theory and practice. Research works dealing with these topics have been presented at ICNLSP 2023.

Thirty seven (37) papers have been accepted by the program committee members that helped us a lot with their insightful comments. All papers have been presented orally, that is why the program was quite long and rich. The technical program included 05 oral sessions, namely: Classification and clustering, Deep learning and transformers, Analysis, summarization, and numerical representation, Speech and phonetics, and Dataset.

This year, we were honoured by the participation of two distinguished scholars: Prof. Dr. Alexander Waibel from Carnegie Mellon University (USA) and Karlsruhe Institute of Technology (Germany) and Dr. Najim Dehak from Johns Hopkins University (USA). Professor Alexander Waibel gave the first talk entitled “Transcending Communication Barriers: From Machine Translation to Language Transparency”. During his talk, Prof. Alex discussed the latest advances and activities to transcend these barriers. The second talk, entitled “Biosignal-based Digital Biomarkers for Aging” was given by Dr. Najim Dehak, in which he described several tools to detect, assess, and monitor the functional and cognitive decline of elderly adults. Both talks were very interesting.

This volume reflects the richness and diversity of the NLP community itself. Contributions from researchers across the globe explore a wide range of languages, domains, and methodologies. This tapestry of research highlights the collaborative spirit and boundless potential of NLP to revolutionize the way we understand, interact with, and create language.

We hope readers enjoy reading the content of the 6th ICNLSP proceedings. We would like also to invite them to check the proceedings of the past versions of ICNLSP.

Mourad Abbas

Organizers:

General Chair: Dr. Mourad Abbas

Chair: Dr. Abed Alhakim Freihat

Program Chair: Dr. Mourad Abbas

Program Committee:

Ahmed Abdelali, SDAIA, Saudi Arabia.

Ahmed Ali, QCRI, Qatar.

Hend Al-Khalifa, King Saud University, Saudi Arabia.

Muhammad Al-Qurishi, Elm, Saudi Arabia.

Mehmet Fatih Amasyalı, Yildiz Technical University, Turkey.

Yuan An, Drexel University, USA.

Fayssal Bouarourou, University of Strasbourg, France.

Hadda Cherroun, Amar Telidji University, Algeria.

Gérard Chollet, CNRS, France.

Dirk Van Compernelle, KU Leuven, Belgium.

Najim Dehak, Johns Hopkins University, USA.

Ashraf Elnagar, University of Sharjah, UAE.

Abed Alhakim Freihat, University of Trento, Italy.

Munir Georges, Technische Hochschule Ingolstadt, Germany.

Ahmed Guessoum, USTHB, Algeria.

Kais Haddar, Faculty of Sciences of Sfax, Tunisia.

Valia Kordoni, Humboldt University, Germany.

Saurabh Kulshreshtha, University of Massachusetts Lowell, USA.

Eric Laporte, Gustave Eiffel University, France.

Mohamed Lichouri, USTHB, Algeria.

Mohammed Mediani, University of Ahmed Draia, Algeria.

Pascal Perrier, Université Grenoble Alpes, France.

Aishwarya Reganti, Amazon Alexa AI.

Hassan Satori, Sidi Mohammed Ben Abdallah University, Morocco.

Tim Schlippe, IU International University of Applied Sciences, Germany.

Nasredine Semmar, CEA, France.

Deven Shah, Microsoft, USA.

R. V. Swaminathan, Amazon, USA.

Irina Temnikova, Big Data for Smart Society Institute, Bulgaria.

Jan Trmal, Johns Hopkins UniIU International University of Applied Sciences, USA.

Fayçal Ykhlef, CDTA, Algeria.

Wajdi Zaghouani, Hamad Bin Khalifa University, Qatar.

Additional reviewers:

Arpit Sood, Meta Inc., USA.

Guanqun Yang, Stevens Institute of Technology, USA.

Abdelouahab Moussaoui, Ferhat Abbas University, Algeria.

Slimane Oulad Naoui, University of Ghardaia, Algeria.

Invited Speakers:

Prof. Dr. Alex Waibel, Carnegie Mellon University, USA.

Prof. Najim Dehak, Johns Hopkins University, USA.

Invited Talks

Transcending Communication Barriers: From Machine Translation to Language Transparency

Prof. Dr. Alex Waibel, Carnegie Mellon University, USA

As we marvel at impressive advances in Artificial Intelligence in recent years, we may wonder whether the problem of language translation and language barriers has been solved. Aside from remaining technical issues, it is important to note that translation is only one (even though important) step toward making people on the planet understand each other: Our thoughts are expressed in many ways: speech, text, video, handwriting, road signs, facial expressions, voice, lip movement, emotion, gesture, mannerisms and more... For frictionless communication, the way technology is deployed in different settings is just as important a consideration as the performance of the technology itself and they come with profound consequences on the technical design and requirements. To make language barriers fade into the background, we need language transparency, not only translation: multimodal, immersive, cross lingual, culturally aware, proactive communication and dubbing tools that interpret the communicative intent and transcend barriers between us. In this talk, I will review major milestones on our journey and discuss our latest advances and activities toward this goal.



Biosignal-based Digital Biomarkers for Aging

Dr. Najim Dehak, Johns Hopkins University, USA

Currently, there are more Americans aged 65 and older (over 49 million) than at any other time in history, according to the US Census Bureau. A significant increase in individuals with severe chronic conditions will have profound social and economic effects on society. Three aspects describe the human aging process: functional (motor system), cognitive, and behavior (social and psychological stressors). In this talk, we will describe several tools to detect, assess, and monitor the functional and cognitive decline of elderly adults. Those tools named biomarkers are based on multimodal biosignals such as speech, handwriting, and eye movement. In addition, we will describe our current work on emotion recognition from speech that can be used to assess social and psychological stressors.

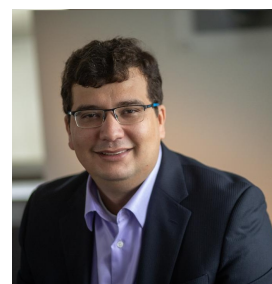


Table of Contents

Classification of Human- and AI-Generated Texts for English, French, German, and Spanish	1
<i>Kristina Schaaff, Tim Schlippe, Lorenz Mindner</i>	
Handling Realistic Label Noise in BERT Text Classification	11
<i>Maha Agro, Hanan Aldarmaki</i>	
Discourse Relations Classification and Cross-Framework Discourse Relation Classification Through the Lens of Cognitive Dimensions: An Empirical Investigation	21
<i>Yingxue Fu</i>	
Representation Learning for Hierarchical Classification of Entity Titles	43
<i>Elena Chistova</i>	
DAP-LeR-DAug: Techniques for enhanced Online Sexism Detection	51
<i>Jayant Panwar, Radhika Mamidi</i>	
CommunityFish: A Poisson-based Document Scaling With Hierarchical Clustering	59
<i>Sami Diaf</i>	
ADCluster: Adaptive Deep Clustering for Unsupervised Learning from Unlabeled Documents	68
<i>Arezoo Hatefi, Xuan-Son Vu, Monowar Bhuyan, Frank Drewes</i>	
Efficient Black-Box Adversarial Attacks on Neural Text Detectors	78
<i>Vitalii Fishchuk, Daniel Braun</i>	
Transformer-Based Analysis of Sentiment Towards German Political Parties on Twitter During the 2021 Election Year	84
<i>Nils Constantin Hellwig, Markus Bink, Thomas Schmidt, Jakob Fehle, Christian Wolff</i>	
"Japan's Answer to Mozart": Automatic Detection of Generalized Patterns of Vossian Antonomasia	99
<i>Michel Schwab, Robert Jäschke, Frank Fischer</i>	
GAVI: A Category-Aware Generative Approach for Brand Value Identification	110
<i>kassem sabeh, Mouna Kacimi, Johann Gamper</i>	
Simple, Simpler and Beyond: A Fine-Tuning BERT-Based Approach to Enhance Sentence Com- plexity Assessment for Text Simplification	120
<i>Lucía Ormaechea, Nikos Tsourakis, Didier Schwab, Pierrette Bouillon, Benjamin Lecouteux</i>	
Deep Learning-Based Claim Matching with Multiple Negatives Training	134
<i>Anna Neumann, Dorothea Kolossa, Robert M Nickel</i>	
Exploring BERT Models for Part-of-Speech Tagging in the Algerian Dialect: A Comprehensive Study	140
<i>Mohamed Amine Cheragui, Abdelhalim Hafedh Dahou, Amin Abdedaiem</i>	
A Neural Network Approach to Ellipsis Detection in Ancient Greek	151
<i>Giuseppe G. A. Celano</i>	
AraBERT and mBert: Insights from Psycholinguistic Diagnostics	159
<i>Basma Sayah, Attia Nehar, Hadda Cherroun, Slimane Bellaouar</i>	
An NLP Analysis of ChatGPT's Personality Simulation Capabilities and Implications for Human- centric Explainable AI Interfaces	168

Thorsten Zylowski, Matthias Wölfel

Topically diversified summarization of customer reviews	178
<i>Florian Carichon, Gilles Caporossi</i>	
Extracting higher-order logic formulas from English sentences	192
<i>Alexandre Rademaker, Guilherme Lima, Renato Cerqueira</i>	
A Quantitative Approach to Understand Self-Supervised Models as Cross-lingual Feature Extracters	200
<i>Shuyue Stella Li, Beining Xu, Xiangyu Zhang, Hexin Liu, Wenhan Chao, Paola Garcia</i>	
Def2Vec: Extensible Word Embeddings from Dictionary Definitions	212
<i>Irene Morazzoni, Vincenzo Scotti, Roberto Tedesco</i>	
Exploring Hybrid Linguistic Features for Turkish Text Readability	223
<i>Ahmet Yavuz Uluslu</i>	
Comparison of Wav2vec 2.0 Transformer Models for Speaker Change Detection	233
<i>Zbyněk Zajíc, Marie Kunešová</i>	
Typological classification of European Portuguese fricatives: a cross-language forced alignment and pronunciation variants study	239
<i>Anisia Popescu, Lori Lamel, Ioana Vasilescu</i>	
Methods for Phonetic Scraping of Youtube Videos	244
<i>Adrien Meli, Steven Coats, Nicolas Ballier</i>	
Direct Speech to Text Translation: Bridging the Modality Gap Using SimSiam	250
<i>Balaram Sarkar, Chandresh K Maurya, Anshuman Agrahri</i>	
Improving Dhivehi Automatic Speech Recognition (ASR) with Sub-word Modelling, Language Model Decoding and Automatic Spelling Correction	256
<i>Arushad Ahmed</i>	
Comparing Modular and End-To-End Approaches in ASR for Well-Resourced and Low-Resourced Languages	266
<i>Aditya Parikh, Louis ten Bosch, Henk van den Heuvel, Cristian Tejedor-Garcia</i>	
Towards Joint Modeling of Dialogue Response and Speech Synthesis based on Large Language Model	274
<i>Xinyu Zhou, DeLong Chen, Yudong Chen</i>	
Using Whisper LLM for Automatic Phonetic Diagnosis of L2 Speech, a Case Study with French Learners of English	282
<i>Nicolas Ballier, Adrien Meli, Maelle Amand, Jean-Baptiste Yunès</i>	
Enhancing Word Discrimination and Matching in Query-by-Example Spoken term detection with Acoustic Word Embeddings	293
<i>Pantid Chantangphol, Theerat Sakdejayont, Tawunrat Chalothorn</i>	
Turkish Native Language Identification	303
<i>Ahmet Yavuz Uluslu</i>	
KMD: A New Kurdish Multilabel Emotional Dataset For the Kurdish Sorani Dialect	308
<i>Soran SM Badawi</i>	

iTANONG-DS : A Collection of Benchmark Datasets for Downstream Natural Language Processing Tasks on Select Philippine Languages	316
<i>Moses L. Visperas, Christalline Joie Borjal, Aunhel John M Adoptante, Danielle Shine R. Abacial, Ma. Miciella Decano, Elmer C Peramo</i>	
Data Augmentation for Text Classification with EASE	324
<i>A M Muntasir Rahman, Wenpeng Yin, Guiling Wang</i>	
Enrichment of Arabic WordNet Using Machine Translation and Transformers	333
<i>Mohamed Dia Eddine Souci, Younes Cherifi, Lamia Berkani, Mohamed Seghir Hadj Ameer, Ahmed Gues-soum</i>	
Compiling a Corpus of Technical Documents for Dialogue System Development in the Industrial Sector	341
<i>Laura García-Sardiña, Eneko Ruiz, Cristina Aceta, Izaskun Fernández, Maria Inés Torres, Arantza del Pozo</i>	