

DAY 2 Oral Sessions			
Session O13: Statistical Methods and Machine Learning (1)	Chair: Vésteinn Snæbjarnarson Co-chair: 57	Auditorium	
9.30-9.50	57	Taja Kuzman, Peter Rupnik and Nikola Ljubešić?	The GINCO Training Dataset for Web Genre Identification of Documents Out in the Wild
9.50-10.10	655	Gaëlle Laperrière, Valentin Pelloin, Antoine Caubrière, salima mdhaffar, Nathalie Camelin, Sahar Ghannay, Bassam Jabaian and Yannick Estève	The Spoken Language Understanding MEDIA Benchmark Dataset in the Era of Deep Learning: data updates, training and evaluation tools
10.10-10.30	820	Gorka Urbizu, Iñaki San Vicente, Xabier Saralegi, Rodrigo Aggerri and Aitor Soroa	BasqueGLUE: A Natural Language Understanding Benchmark for Basque
10.30-10.50	980	Nicolas Stefanovitch, Jakub Piskorski and Sopho Kharazi	Resources and Experiments on Sentiment Classification for Georgian
Session O14: Natural Language Generation and Summarisation	Chair: Inguna Skadina Co-chair: Kelvin Han	La Major	
9.30-9.50	178	Nadhem ZMANDAR, Tobias Daudert, Sina Ahmadi, Mahmoud El-Haj and Paul Rayson	CoFIF Plus: A French Financial Narrative Summarisation Corpus
9.50-10.10	550	Rémi Calizzano, Malte Ostendorff, Qian Ruan and Georg Rehm	Generating Extended and Multilingual Summaries with Pre-trained Transformers
10.10-10.30	903	Louis Raphaël Théo Martin, Angela Fan, Éric de la Clergerie, Antoine Bordes and Benoît Sagot	MUSS: Multilingual Unsupervised Sentence Simplification by Mining Paraphrases
10.30-10.50	1130	Samhita Honnavalli, Aesha Parekh, Lily Ou, Sophie Groenwold, Sharon Levy, Vicente Ordóñez and William Yang Wang	Towards Understanding Gender-Seniority Compound Bias in Natural Language Generation
Session O15: Semantics	Chair: Bolette Pedersen Co-chair: Mehwish Alam	Salle 120	
9.30-9.50	492	Fabio Tamburini	Combining ELECTRA and Adaptive Graph Encoding for Frame Identification
9.50-10.10	766	Aina Garí Soler, Matthieu Labeau and Chloé Clavel	Polysemy in Spoken Conversations and Written Texts
10.10-10.30	938	Vuk Batanovic and Maja Milicevic Petrovic	Cross-Level Semantic Similarity for Serbian Newswire Texts
10.30-10.50	1089	Ishan Jindal, Alexandre Rademaker, Micha? Ulewicz, Ha Linh, Huyen Nguyen, Khoi-Nguyen Tran, Huaiyu Zhu and Yunyao Li	Universal Proposition Bank 2.0
Session O16: Language Resources and Evaluation for Psycholinguistics and Cognitive Linguistics	Chair: Patrizia Paggio Co-chair: Harish Tayyar Madabushi	Salle 92	
9.30-9.50	775	Nora Hollenstein, Maria Barrett and Marina Björnsdóttir	The Copenhagen Corpus of Eye Tracking Recordings from Natural Reading of Danish Texts
9.50-10.10	1046	Andreas Weise, Matthew McNeill and Rivka Levitan	The Brooklyn Multi-Interaction Corpus for Analyzing Variation in Entrainment Behavior
10.10-10.30	779	Aleksandra Miletic, Christophe Benzitoun, Georgeta Cislaru and Santiago Herrera-Yanez	Pro-TEXT: an Annotated Corpus of Keystroke Logs
10.30-10.50	392	Federico Bonetti, Elisa Leonardelli, Daniela Trotta, Raffaele Guarasci, Sara Tonelli	Work Hard, Play Hard: Collecting Acceptability Annotations through a 3D Game
<i>10.50-11.10 Coffee Break</i>			
Session O17: Evaluation and Validation Methodologies (1)	Chair: Menno van Zaanen Co-chair: Maria Eskevich	Auditorium	
11.10-11.30	960	Rinalds Viksna, Inguna Skadina, Raivis Skadinš, Andrejs Vasiljevs and Roberts Rozis	Assessing Multilinguality of Publicly Accessible Websites
11.30-11.50	535	David Kletz, Philippe Langlais, François Lareau and Patrick Drouin	A Methodology for Building a Diachronic Dataset of Semantic Shifts and its Application to QC-FR-Diac-V1.0, a Free Reference for French
11.50-12.10	134	Jörg Froberg and Frank Binder	CRASS: A Novel Data Set and Benchmark to Test Counterfactual Reasoning of Large Language Models
12.10-12.30	40	Marta R. Costa-jussà, Christine Basta and Gerard I. Gállego	Evaluating Gender Bias in Speech Translation
Session O18: Applications Involving LRs and Evaluation (2)	Chair: Yoshihiko Hayashi Co-chair: Maria Barrett	La Major	
11.10-11.30	865	Merel Scholman, Valentina Pyatkin, Frances Yung, Ido Dagan, Reut Tsarfaty and Vera Demberg	Design Choices in Crowdsourcing Discourse Relation Annotations: The Effect of Worker Selection and Training
11.30-11.50	998	Hrishikesh Kulkarni, Sean MacAvaney, Nazli Goharian and Ophir Frieder	TBD3: A Thresholding-Based Dynamic Depression Detection from Social Media for Low-Resource Users
11.50-12.10	1017	Sayontan Ghosh, Amanpreet Singh, Alex Merenstein, Wei Su, Scott A. Smolka, Erez Zadok and Niranjana Balasubramanian	SpecNFS: A Challenge Dataset Towards Extracting Formal Models from Natural Language Specifications
12.10-12.30	368	Xiaoyu Bai, Manfred Stede	Argument Similarity Assessment in German for Intelligent Tutoring: Crowdsourced Dataset and First Experiments
Session O19: Information Extraction & Neural Networks	Chair: Jan Hajič Co-chair: Atul Kumar Ojha	Salle 120	
11.10-11.30	1267	Nishtha Jain, Declan Groves, Lucia Specia and Maja Popovic	Leveraging Pre-trained Language Models for Gender Debiasing
11.30-11.50	44	Gretel Liz De la Peña Sarracén and Paolo Rosso	Unsupervised Embeddings with Graph Auto-Encoders for Multi-domain and Multilingual Hate Speech Detection
11.50-12.10	229	Quentin Heinrich, Gautier Viaud and Wacim Belbidia	FQuAD2.0: French Question Answering and Learning When You Don't Know
12.10-12.30	303	Cagri Toraman, Furkan Şahinuç and Eyup Halit Yilmaz	Large-Scale Hate Speech Detection with Cross-Domain Transfer
Session O20: Dialogue	Chair: Mathilde Hutin Co-chair: 728	Salle 92	
11.10-11.30	728	Selina Meyer and David Elsweiler	GLoHBCD: A Naturalistic German Dataset for Language of Health Behaviour Change on Online Support Forums
11.30-11.50	848	Virginia Meijer and Iris Hendrickx	Creating a Data Set of Abstractive Summaries of Turn-labeled Spoken Human-Computer Conversations
11.50-12.10	1051	Wen Cui, Leanne Rolston, Marilyn Walker and Beth Ann Hockey	OpenEL: An Annotated Corpus for Entity Linking and Discourse in Open Domain Dialogue
12.10-12.30	1246	Bram Willemsen, Dmytro Kalpakchi, Gabriel Skantze	Collecting Visually-grounded Argumentative Dialogue with A Game Of Sorts
12.30-13.00 Invited Local Talk - Philippe Boula de Mareuil	Chair: Philippe Blache	Auditorium	The languages of France illustrated by a Speaking Atlas
<i>13.00-14.30 Lunch Break</i>			
14.30-15.10 Keynote Speaker - Emmanuel Dupoux	Chair: Joseph Mariani	Auditorium	Textless NLP: towards language processing from raw audio
<i>15.10-15.15 5-minutes break</i>			
Session O21: Corpora and Annotation (2)	Chair: Gilles Adda Co-chair: Valeria Quochi	Auditorium	
15.15-15.35	674	Jan Philip Wahle, Terry Ruas, Saif M. Mohammad and Bela Gipp	D3: A Massive Dataset of Scholarly Metadata for Analyzing the State of Computer Science Research
15.35-15.55	683	Dimitrios Roussis, Vassilis Papavassiliou, Prokopis Prokopidis, Stelios Piperidis and Vassilis Katsouras	SciPar: A Collection of Parallel Corpora from Scientific Abstracts
15.55-16.15	893	Martha Gavidia, Patrick Lee, Anna Feldman and Jing Peng	CATs are Fuzzy PETS: A Corpus and Analysis of Potentially Euphemistic Terms
16.15-16.35	1195	Nizar Habash, Muhammed AbuOdeh, Dima Taji, Reem Faraj, Jamila El Gizuli and Omar Kallas	Camel Treebank: An Open Multi-genre Arabic Dependency Treebank

Session 022: Summarization	Chair: Fabio Tamburini Co-chair:	La Major	
15.15-15.35	928	Sajad Sotudeh, Nazli Goharian and Zachary Young	MentSum: A Resource for Exploring Summarization of Mental Health Online Posts
15.35-15.55	1118	Dennis Aumiller and Michael Gertz	Klexikon: A German Dataset for Joint Summarization and Simplification
15.55-16.15	982	Philipp Hartl and Udo Kruschwitz	Applying Automatic Text Summarization for Fake News Detection
16.15-16.35			
Session 023: Language Resource Infrastructures and Standards	Chair: Franciska De Jong Co-chair: Kaja Dobrovoljc	Salle 120	
15.15-15.35	341	Nino Meisinger, Thorsten Trippel and Claus Zinn	Increasing CMDI's Semantic Interoperability with schema.org
15.35-15.55	1005	Herbert Lange and Jocelyn Aznar	RefCo and Its Checker: Improving Language Documentation Corpora's Reusability Through a Semi-Automatic Review Process
15.55-16.15	203	Gábor Simon	Identification and Analysis of Personification in Hungarian: The PerSECorp project
16.15-16.35	926	Purificação Moura Silvano, Mariana K. Damova, Giedrė Valūnaite Oleškevičienė, Chaya Liebeskind, Christian Chiarcos, Dimitar Trajanov, Ciprian-Octavian Truiță, Elena-Simona Apostol and Anna Baczkowska	ISO based Discourse Markers Annotated Multilingual Corpus
Session 024: Multimodality and Cross-modality	Chair: Eva Hajicova Co-chair: Aina Gari Soler	Salle 92	
15.15-15.35	69	David Gimeno-Gómez and Carlos-D. Martínez-Hinarejos	LIP-RTVE: An Audiovisual Database for Continuous Spanish in the Wild
15.35-15.55	239	Hyeonju Yun, Yongil Kim and Kyomin Jung	Modality Alignment between Deep Representations for Effective Video-and-Language Learning
15.55-16.15	398	Anais Murat, Maria Koutsombogera and Carl Vogel	Mutual Gaze and Linguistic Repetition in a Multimodal Corpus
16.15-16.35	774	Christophe Parisse, Marion Blondel, Stéphanie Caët, Claire Danet, Coralie Vincent and Aliyah Morgenstern	Multidimensional Coding of Multimodal Language in Multi-Party Settings
16.35-16.55 <i>Coffee Break</i>			
Session 025: Social Media Processing	Chair: Paul Rayson Co-chair: Itziar Aldabe	Auditorium	
16.55-17.15	298	Kathleen C. Fraser, Svetlana Kiritchenko and Isar Nejadgholi	Extracting Age-Related Stereotypes from Social Media Texts
17.15-17.35	723	Elena Alvarez-Mellado and Constantine Lignos	Borrowing or Codeswitching? Annotating for Finer-Grained Distinctions in Language Mixing
17.35-17.55	764	Ana Sabina Uban, Berta Chulvi and Paolo Rosso	Multi-Aspect Transfer Learning for Detecting Low Resource Mental Disorders on Social Media
17.55-18.15	784	Hamdy Mubarak, Sabit Hassan, Shammur Absar Chowdhury and Firoj Alam	ArCovidVac: Analyzing Arabic Tweets About COVID-19 Vaccination
18.15-18.35	1194	Flora Sakketou, Joan Plepi, Riccardo Cervero, Henri Jacques Geiss, Paolo Rosso and Lucie Flek	FACTOID: A New Dataset for Identifying Misinformation Spreaders and Political Bias
Session 026: Speech Resources and Processing	Chair: Krister Lindén Co-chair: Chiara Mazzocconi	La Major	
16.55-17.15	50	Julia Maria Pritzen, Michael Gref, Dietlind Zühlke and Christoph Andreas Schmidt	Multitask Learning for Grapheme-to-Phoneme Conversion of Anglicisms in German Speech Recognition
17.15-17.35	425	Michel Plüss, Manuela Hürlimann, Marc André Cuny, Alla Stöckli, Nikolaos Kapatzi, Julia Hartmann, Malgorzata Anna Ulasik, Christian Scheller, Yanick Schraner, Amit Jain, Jan Deriu, Mark Cieliebak and Manfred Vogel	SDS-200: A Swiss German Speech to Standard German Text Corpus
17.35-17.55	555	Yaru WU, Mathilde Hutin, Ioana Vasilescu, Lori Lamel and Martine Adda-Decker	Extracting Linguistic Knowledge from Speech: A Study of Stop Realization in 5 Romance Languages
17.55-18.15	701	Martin Lebourdais, Marie Tahon, Antoine LAURENT, Sylvain Meignier and Anthony Larcher	Study of Overlaps and Gender Annotations in the Context of Broadcast Media
18.15-18.35	896	Rémi Uro, David Doukhan, Albert Rilliard, Laetitia Larcher, Anissa-Claire Adgharouamane, Marie Tahon and Antoine Laurent	A Semi-Automatic Approach to Create Large Gender- and Age-Balanced Speaker Corpora: Usefulness of Speaker Diarization & Identification
Session 027: Discourse	Chair: Elena Cabrio Co-chair: Gavin Abercrombie	Salle 120	
16.55-17.15	484	Merel Scholman, Tianai Dong, Frances Yung and Vera Demberg	DiscoGeM: A Crowdsourced Corpus of Genre-Mixed Implicit Discourse Relations
17.15-17.35	605	Annette Hautli-Janisz, Zlata Kikteva, Wassiliki Siskou, Kamila Gorska, Ray Becker and Chris Reed	QT30: A Corpus of Argument and Conflict in Broadcast Debate
17.35-17.55	643	Neele Falk and Gabriella Lapesa	Scaling up Discourse Quality Annotation for Political Science
17.55-18.15	702	Talita Anthonio, Anna Sauer and Michael Roth	Clarifying Implicit and Underspecified Phrases in Instructional Text
18.15-18.35	1218	Anton Buzanov, Polina Bychkova, Arina Molchanova, Anna Postnikova and Daria Ryzhova	Multilingual Pragmaticon: Database of Discourse Formulae
Session 028: Digital Humanities and Cultural Heritage	Chair: Andreas Witt Co-chair: Simona Frenda	Salle 92	
16.55-17.15	554	Ranka Stanković, Cvetana Krstev, Branislava Šandrih Todorović, Dusko Vitas, Mihailo Skoric and Milica Ikončić Nešić	Distant Reading in Digital Humanities: Case Study on the Serbian Part of the ELTeC Collection
17.15-17.35	806	Nils Reiter, Judith Sieker, Svenja Guhr, Evelyn Gius and Sina Zarriß	Exploring Text Recombination for Automatic Narrative Level Detection
17.35-17.55	873	Rachel Bawden, Jonathan Poinhos, Eleni Kogkitsidou, Philippe Gambette, Benoît Sagot and Simon Gabay	Automatic Normalisation of Early Modern French
17.55-18.15	875	Simon Gabay, Pedro Javier Ortiz Suarez, Alexandre BARTZ, Alix Chagué, Rachel Bawden, Philippe Gambette and Benoît Sagot	From FreEM to D'AleMBERT: a Large Corpus and a Language Model for Early Modern French
18.15-18.35	990	Nuette Heyns and Menno van Zaanen	Detecting Multiple Transitions in Literary Texts