

CLiC-it 2021

**The Eighth Italian Conference on Computational Linguistics**

**Proceedings of the Conference**

June 29 - July 1, 2022

Copyright ©2021 for the individual papers by the papers' authors.  
Copyright ©2021 for the volume as a collection by its editors.  
This volume and its papers are published under the Creative Commons License Attribution 4.0 International (CC BY 4.0).

These proceedings have been published in the CEUR Workshop Proceedings series.  
The original papers are available at: <https://ceur-ws.org/Vol-3033>.  
The papers are mirrored in the ACL Anthology.

ISBN 979-12-80136-94-7

## Table of Contents

<i>Preface</i>	
Elisabetta Fersini, Marco Passarotti and Viviana Patti . . . . .	1
<i>Returning the L in NLP: Why Language (Variety) Matters and How to Embrace it in Our Models (key-note)</i>	
Barbara Plank . . . . .	7
<i>It's Time for Reasoning (keynote)</i>	
Dan Roth . . . . .	9
<i>Introduction to Linguistic Linked Open Data (tutorial)</i>	
Christian Chiarcos . . . . .	10
<i>Universal Dependencies: Principles and Tools (tutorial)</i>	
Daniel Zeman . . . . .	12
<i>Atypical or Underrepresented? A Pilot Study on Small Treebanks</i>	
Akshay Aggarwal and Chiara Alzetta . . . . .	13
<i>On the Role of Textual Connectives in Sentence Comprehension: A New Dataset for Italian</i>	
Giorgia Albertin, Alessio Miaschi and Dominique Brunato . . . . .	22
<i>Query in linguaggio naturale per il dominio della dieta mediterranea(Natural Language Queries for the Mediterranean Diet Domain)</i>	
Luca Anselma, Dario Ferrero and Alessandro Mazzei . . . . .	30
<i>Are Crescia and Piadina the Same? Towards Identifying Synonymy or Non-Synonymy between Italian Words to Enable Crowdsourcing from Language Learners</i>	
Lavinia Aparaschivei, Lionel Nicolas and Alberto Barrón-Cedeño . . . . .	37
<i>Moving from Human Ratings to Word Vectors to Classify People With Focal Dementias: Are We There Yet?</i>	
Chiara Barattieri Di San Pietro, Marco Marelli and Carlo Reverberi . . . . .	44
<i>WITS: Wikipedia for Italian Text Summarization</i>	
Silvia Casola and Alberto Lavelli . . . . .	51
<i>Polycorpus XL: An Italian Corpus for the Detection of Hate Speech Against Politics</i>	
Fabio Celli, Mirko Lai, Armend Duzha, Cristina Bosco and Viviana Patti . . . . .	58
<i>Quale testo è scritto meglio? A Study on Italian Native Speakers' Perception of Writing Quality</i>	
Aldo Cerulli, Dominique Brunato and Felice Dell'Orletta . . . . .	65
<i>On the Development of Customized Neural Machine Translation Models</i>	
Mauro Cettolo, Roldano Cattoni and Marco Turchi . . . . .	72
<i>A Common Derivation for Parsing and Generation With Expectation-Based Minimalist Grammars (e-MGs)</i>	
Cristiano Chesi . . . . .	79
<i>La ministro è incinta": A Twitter Account of Women's Job Titles in Italian</i>	
Alessandra Teresa Cignarella, Mirko Lai, Andrea Marra and Manuela Sanguinetti . . . . .	87

<i>GQA-it: Italian Question Answering on Image Scene Graphs</i> Danilo Croce, Lucia C. Passaro, Alessandro Lenci and Roberto Basili .....	94
<i>Implementing a Pragmatically Adequate Chatbot in DialogFlow CX</i> Anna Dall'Acqua and Fabio Tamburini .....	102
<i>La produzione di testi semplificati di notiziari televisivi italiani destinati a persone con disturbi cognitivi acquisiti: un'integrazione tra metodi psicolinguistici e analisi automatiche(Implementing Simplified TV News Texts in Italian for People with Acquired Cognitive Disorders: Psycholinguistic Methods and Automatic Analyses)</i> Maria De Martino and Annalisa Colella .....	110
<i>Dialogue Analysis with Graph Databases: Characterising Domain Items Usage for Movie Recommendations</i> Martina Di Bratto, Maria Di Maro, Antonio Origlia and Francesco Cutugno .....	116
<i>Do You Have Any Recommendation? An Annotation System For The Seekers' Strategies In Recommendation Dialogues</i> Martina Di Bratto, Riccardo Orrico, Ancuta Budeanu, Marta Maffia and Loredana Schettino	123
<i>Language Transfer for Identifying Diagnostic Paragraphs in Clinical Notes</i> Luca Di Liello, Olga Uryupina and Alessandro Moschitti .....	130
<i>Trattamento automatico della lingua a supporto dell'editoria: primi esperimenti con il Devoto-Oli Junior(Automatic Language Treatment to Support Publishing: First Experiments with the Devoto-Oli Junior)</i> Irene Dini, Felice Dell'Orletta, Fabio Ferri, Biancamaria Gismondi and Simonetta Montemagni	137
<i>A First Step Towards Automatic Consolidation of Legal Acts: Reliable Classification of Textual Modifications</i> Samuel Fabrizi, Maria Iacono, Andrea Tesei and Lorenzo De Mattei .....	143
<i>Hate Speech and Topic Shift in the Covid-19 Public Discourse on Social Media in Italy</i> Komal Florio, Valerio Basile and Viviana Patti .....	150
<i>Recognizing Hate with NLP: The Teaching Experience of the #DeactivHate Lab in Italian High Schools</i> Simona Frenda, Alessandra Teresa Cignarella, Marco Antonio Stranisci, Mirko Lai, Cristina Bosco and Viviana Patti .....	157
<i>The Role of a Computational Lexicon for Query Expansion in Full-Text Search</i> Emiliano Giovannetti, Davide Albanesi, Andrea Bellandi, Simone Marchi, Mafalda Papini and Flavia Sciolette .....	164
<i>A Methodology for Large-Scale, Disambiguated and Unbiased Lexical Knowledge Acquisition Based on Multilingual Word Alignment</i> Francesca Grasso and Luigi Di Caro .....	171
<i>The Annotation of Liber Abbaci, a Domain-Specific Latin Resource</i> Francesco Grotto, Rachele Sprugnoli, Margherita Fantoli, Maria Simi, Flavio Massimiliano Cecchini and Marco Passarotti .....	178
<i>Detecting Age-Related Linguistic Patterns in Dialogue: Toward Adaptive Conversational Systems</i> Lennert Jansen, Arabella Sinclair, Margot J. van der Goot, Raquel Fernández and Sandro Pezzelle	186

<i>From Cambridge to Pisa: A Journey into Cross-Lingual Dialogue Domain Adaptation for Conversational Agents</i>	
Tiziano Labruna and Bernardo Magnini . . . . .	193
<i>Agentività e telicità in GILBERTo: implicazioni cognitive (Agentivity and Telicity in GILBERTo: Cognitive Implications)</i>	
Agnese Lombardi and Alessandro Lenci . . . . .	200
<i>Investigating Continued Pretraining for Zero-Shot Cross-Lingual Spoken Language Understanding</i>	
Samuel Louvan, Silvia Casola and Bernardo Magnini . . . . .	207
<i>Linking the Lewis &amp; Short Dictionary to the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin</i>	
Francesco Mambrini, Eleonora Litta, Marco Passarotti and Paolo Ruffolo . . . . .	216
<i>The ListTyp Database</i>	
Francesca Masini, Simone Mattiola and Stefano Dei Rossi . . . . .	223
<i>Probing Tasks Under Pressure</i>	
Alessio Miaschi, Chiara Alzetta, Dominique Brunato, Felice Dell’Orletta and Giulia Venturi	230
<i>Frame Semantics for Social NLP in Italian: Analyzing Responsibility Framing in Femicide News Reports</i>	
Gosse Minnema, Sara Gemelli, Chiara Zanchi, Viviana Patti, Tommaso Caselli and Malvina Nissim . . . . .	237
<i>Leveraging Bias in Pre-Trained Word Embeddings for Unsupervised Microaggression Detection</i>	
Tolúlope Ògúnremí, Nazanin Sabri, Valerio Basile and Tommaso Caselli . . . . .	245
<i>REDIT: A Tool and Dataset for Extraction of Personal Data in Documents of the Public Administration Domain</i>	
Teresa Paccosi and Alessio Palmero Aprosio . . . . .	252
<i>It Is MarkIT That Is New: An Italian Treebank of Marked Constructions</i>	
Teresa Paccosi, Alessio Palmero Aprosio and Sara Tonelli . . . . .	260
<i>Linguistic Cues of Deception in a Multilingual April Fools’ Day Context</i>	
Katerina Papantoniou, Panagiotis Papadakos, Giorgos Flouris and Dimitris Plexousakis . . . . .	267
<i>Visualization: The Missing Factor in Simultaneous Speech Translation</i>	
Sara Papi, Matteo Negri and Marco Turchi . . . . .	274
<i>Audience Engagement Prediction in Guided Tours through Multimodal Features</i>	
Andrea Amelio Ravelli, Andrea Cimino and Felice Dell’Orletta . . . . .	282
<i>FANCY: A Diagnostic Data-Set for NLI Models</i>	
Guido Rocchietti, Flavia Achena, Giuseppe Marziano, Sara Salaris and Alessandro Lenci . . . . .	289
<i>Automatic Assessment of English CEFR Levels Using BERT Embeddings</i>	
Veronica Juliana Schmalz and Alessio Brutti . . . . .	295
<i>Introducing a Gold Standard Corpus from Young Multilinguals for the Evaluation of Automatic UD-PoS Taggers for Italian</i>	
Veronica Juliana Schmalz, Jennifer-Carmen Frey and Egon W. Stemle . . . . .	302
<i>Extracting Relations from Italian Wikipedia using Self-Training</i>	
Lucia Siciliani, Pierluigi Cassotti, Pierpaolo Basile, Marco de Gemmis, Pasquale Lops and Giovanni Semeraro . . . . .	309

<i>Sentiment Analysis of Latin Poetry: First Experiments on the Odes of Horace</i> Rachele Sprugnoli, Francesco Mambrini, Marco Passarotti and Giovanni Moretti . . . . .	316
<i>Playing with NeMo for Building an Automatic Speech Recogniser for Italian</i> Fabio Tamburini . . . . .	323
<i>OCTIS 2.0: Optimizing and Comparing Topic Models in Italian Is Even Simpler!</i> Silvia Terragni and Elisabetta Fersini . . . . .	330
<i>How Contextualized Word Embeddings Represent Word Senses</i> Rocco Tripodi . . . . .	337
<i>ConteCorpus: An Analysis of People Response to Institutional Communications During the Pandemic</i> Viviana Ventura and Elisabetta Jezek . . . . .	346
<i>PROTECT - A Pipeline for Propaganda Detection and Classification</i> Vorakit Vorakitphan, Elena Cabrio and Serena Villata . . . . .	354
<i>A Multi-Strategy Approach to Crossword Clue Answer Retrieval and Ranking</i> Andrea Zugarini and Marco Ernandes . . . . .	361
<i>Emerging Trends in Gender-Specific Occupational Titles in Italian Newspapers</i> Pierluigi Cassotti, Andrea Iovine, Pierpaolo Basile, Marco De Gemmis and Giovanni Semeraro 368	
<i>Deep Learning Representations in Automatic Misogyny Identification: What Do We Gain and What Do We Miss?</i> Elisabetta Fersini, Luca Rosato, Antonio Candelieri, Francesco Archetti and Enza Messina . .	374
<i>An Obligations Extraction System for Heterogeneous Legal Documents: Building and Evaluating Data and Model</i> Maria Iacono, Laura Rossi, Paolo Dangelo, Andrea Tesei and Lorenzo De Mattei . . . . .	380
<i>FONTI 4.0: Evaluating Speech-to-Text Automatic Transcription of Digitized Historical Oral Sources</i> Roberta Bianca Luzietti, Niccolò Pretto, Frédéric Kaplan, Alain Dufaux and Sergio Canazza	386
<i>An Unsupervised Approach to Extract Life-Events from Personal Narratives in the Mental Health Domain</i> Seyed Mahed Mousavi, Roberto Negro and Giuseppe Riccardi . . . . .	393
<i>An Italian Question Answering System Based on Grammars Automatically Generated from Ontology Lexica</i> Gennaro Nolano, Mohammad Fazleh Elahi, Maria Pia Di Buono, Basil Ell and Philipp Cimiano 398	
<i>Tackling Italian University Assessment Tests with Transformer-Based Language Models</i> Daniele Puccinelli, Silvia Demartini and Pier Luigi Ferrari . . . . .	404
<i>KERMIT for Sentiment Analysis in Italian Healthcare Reviews</i> Leonardo Ranaldi, Michele Mastromattei, Dario Onorati, Elena Sofia Ruzzetti, Francesca Fallucchi and Fabio Massimo Zanzotto . . . . .	410
<i>T-PAS Scraper: An Application for Linguistic Data Extraction and Analysis</i> Emma Romani, Valerio Gattero and Elisabetta Jezek . . . . .	416