

CLiC-it 2023

The Ninth Italian Conference on Computational Linguistics

Proceedings of the Conference

November 30 - December 2, 2023

Copyright ©2023 for the individual papers by the papers' authors.
Copyright ©2023 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-3596>.
The papers are mirrored in the ACL Anthology.

ISBN 979-12-550-0084-6

Table of Contents

<i>Preface to the CLiC-it 2023 Proceedings</i>	
Federico Boschetti, Gianluca E. Lebani, Bernardo Magnini and Nicole Novielli	1
<i>When the Lab of CompLing Was Started at the University of Venice - Preface to the Proceedings of the First Workshop Held in 1982</i>	
Rodolfo Delmonte	4
<i>Legal Summarization: to Each Court its Own Model</i>	
Flavia Aचना, David Preti, Davide Venditti, Leonardo Ranaldi, Cristina Giannone, Fabio Massimo Zanzotto, Andrea Favalli and Raniero Romagnoli	10
<i>CorpusCompass: A Tool for Data Extraction and Dataset Generation in Corpus Linguistics</i>	
Muhadj Adnan and Nicolo' Brandizzi	17
<i>Harnessing 'Il Manifesto' Newspaper Archive for Knowledge Base Creation: Techniques and Findings in the 'MeMa' Project</i>	
Robert J. Alexander, Matteo Bartocci, Oriana Persico and Guido Vetere	28
<i>Unmasking the Wordsmith: Revealing Author Identity through Reader Reviews</i>	
Chiara Alzetta, Felice Dell'Orletta, Chiara Fazzone, Alessio Miaschi and Giulia Venturi	36
<i>Integrated Gradients as Proxy of Disagreement in Hateful Content</i>	
Alessandro Astorino, Giulia Rizzi and Elisabetta Fersini	46
<i>Challenging Specialized Transformers on Zero-shot Classification</i>	
Serena Auriemma, Mauro Madeddu, Martina Miliani, Alessandro Bondielli, Alessandro Lenci and Lucia Passaro	53
<i>Bias Mitigation in Misogynous Meme Recognition: A Preliminary Study</i>	
Gianmaria Balducci, Giulia Rizzi and Elisabetta Fersini	63
<i>Building a Spoken Dialogue System for Supporting Blind People in Accessing Mathematical Expressions</i>	
Pier Felice Balestrucci, Luca Anselma, Cristian Bernareggi and Alessandro Mazzei	70
<i>Contrastive Language-Image Pre-training for the Italian Language</i>	
Federico Bianchi, Giuseppe Attanasio, Raphael Pisoni, Silvia Terragni, Gabriele Sarti and Dario Balestri	78
<i>Modelling and Publishing the Lexicon der indogermanischen Verbenas Linked Open Data</i>	
Valeria Irene Boano, Francesco Mambrini, Marco Passarotti and Riccardo Ginevra	86
<i>How Green is Sentiment Analysis? Environmental Topics in Corpora at the University of Turin</i>	
Cristina Bosco, Muhammad Okky Ibrohim, Valerio Basile and Indra Budi	93
<i>Ti blocco perché sei un trollazzo". Lexical Innovation in Contemporary Italian in a Large Twitter Corpus</i>	
Paolo Brasolin, Greta Hayley Franzini and Stefania Spina	100
<i>Testing ChatGPT for Stability and Reasoning: A Case Study Using Italian Medical Specialty Tests</i>	
Silvia Casola, Tiziano Labruna, Alberto Lavelli and Bernardo Magnini	113
<i>Assessing Language and Vision-Language Models on Event Plausibility</i>	
Maria Cassese, Alessandro Bondielli and Alessandro Lenci	120

<i>GPT-based Language Models meet Emojitaliano: A Preliminary Assessment Test between Automation and Creativity</i>	
Francesca Chiusaroli, Tiberio Uricchio, Johanna Monti, Maria Laura Pierucci and Federico Sangati	128
<i>Linking Stance and Stereotypes About Migrants in Italian Fake News</i>	
Alessandra Teresa Cignarella, Simona Frenda, Tom Bourgeade, Cristina Bosco and Francesca D’Errico	135
<i>Interpretation of Generalization in Masked Language Models: An Investigation Straddling Quantifiers and Generics</i>	
Claudia Collacciani and Giulia Rambelli	143
<i>Highway to Hell. Towards a Universal Dependencies Treebank for Dante Alighieri’s Comedy</i>	
Claudia Corbetta, Marco Passarotti, Flavio Massimiliano Cecchini and Giovanni Moretti	154
<i>Towards an Italian Corpus for Implicit Object Completion</i>	
Agnese Daffara and Elisabetta Jezek	162
<i>Linking the Corpus CLaSSES to the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin</i>	
Irene De Felice, Lucia Tamponi, Federica Iurescia and Marco Passarotti	172
<i>Processing Effort During Reading Texts in Young Adults: Text Simplification, Readability Assessment and Preliminary Eye-tracking Data</i>	
Maria De Martino	179
<i>An Experiment in Error Analysis of Real-time Speech Machine Translation Using the Example of the European Parliament’s Innovation Partnership</i>	
Elisa Di Nuovo	185
<i>A Cognitive Linguistics Analysis of Phrasal Verbs Representation in Distributional Semantics</i>	
Melissa Donati and Carlo Strapparava	196
<i>How To Build Competitive Multi-gender Speech Translation Models For Controlling Speaker Gender Translation</i>	
Marco Gaido, Dennis Fucci, Matteo Negri and Luisa Bentivogli	203
<i>Hate Speech Detection in an Italian Incel Forum Using Bilingual Data for Pre-Training and Fine-Tuning</i>	
Paolo Gajo, Silvia Bernardini, Adriano Ferraresi and Alberto Barrón-Cedeño	211
<i>Linking the Dictionary of Medieval Latin in the Czech Lands to the LiLa Knowledge Base</i>	
Federica Gamba, Marco C. Passarotti and Paolo Ruffolo	219
<i>Check-IT!: A Corpus of Expert Fact-checked Claims for Italian</i>	
Jacopo Gili, Lucia Passaro and Tommaso Caselli	227
<i>End-to-end Dependency Parsing via Auto-regressive Large Language Model</i>	
Claudiu Daniel Hromei, Danilo Croce and Roberto Basili	236
<i>DisaggregHate It Corpus: A Disaggregated Italian Dataset of Hate Speech</i>	
Marco Madeddu, Simona Frenda, Mirko Lai, Viviana Patti and Valerio Basile	243
<i>Automatic Detection of Parkinson’s Disease with Connected Speech Acoustic Features: Towards a Linguistically Interpretable Approach</i>	
Marta Maffia, Loredana Schettino and Vincenzo Norman Vitale	251

<i>Introducing Deep Learning with Data Augmentation and Corpus Construction for LIS</i> Manuela Marchisio, Alessandro Mazzei and Dario Sammaruga	259
<i>XL-WA: a Gold Evaluation Benchmark for Word Alignment in 14 Language Pairs</i> Federico Martelli, Andrei Stefan Bejgu, Cesare Campagnano, Jaka Čibej, Rute Costa, Apolonija Gantar, Jelena Kallas, Svetla Peneva Koeva, Kristina Koppel, Simon Krek, Margit Langemets, Veronika Lipp, Sanni Nimb, Sussi Olsen, Bolette Sanford Pedersen, Valeria Quochi, Ana Salgado, László Simon, Carole Tiberius, Rafael-J Ureña-Ruiz and Roberto Navigli	272
<i>Is Change the Only Constant? An Inquiry into Diachronic Semantic Shifts in Italian and Spanish</i> Matteo Melis, Anastasiia Salova and Roberto Zamparelli	281
<i>Building Structured Synthetic Datasets: The Case of Blackbird Language Matrices (BLMs)</i> Paola Merlo, Giuseppe Samo, Vivi Nastase and Chunyang Jiang	292
<i>Simplifying Administrative Texts for Italian L2 Readers with Controllable Transformers Models: A Data-driven Approach</i> Martina Miliani, Fernando Alva-Manchego and Alessandro Lenci	303
<i>Multi-task Learning for German Text Readability Assessment</i> Salar Mohtaj, Vera Schmitt, Razieh Khamsehashari and Sebastian Möller	316
<i>Is It Really That Simple? Prompting Large Language Models for Automatic Text Simplification in Italian</i> Debora Nozza and Giuseppe Attanasio	322
<i>Unraveling Text Coherence from the Human Perspective: a Novel Dataset for Italian</i> Federica Papa, Luca Dini, Dominique Brunato and Felice Dell’Orletta	334
<i>Lost in Labels: An Ongoing Quest to Optimize Text-to-Text Label Selection for Classification</i> Michele Papucci, Alessio Miaschi and Felice Dell’Orletta	342
<i>Are All Languages Equal? Curriculum Learning over Different Languages</i> Giulia Pucci, Leonardo Ranaldi and Fabio Massimo Zanzotto	351
<i>Prompting LLMs in Italian Language for Text-to-SQL Translation</i> Federico Ranaldi, Elena Sofia Ruzzetti, Leonardo Ranaldi, Davide Venditti, Cristina Giannone, Andrea Favalli, Raniero Romagnoli and Fabio Massimo Zanzotto	361
<i>Beyond Obscuration and Visibility: Thoughts on the Different Strategies of Gender-Fair Language in Italian</i> Martina Rosola, Simona Frenda, Alessandra Teresa Cignarella, Matteo Pellegrini, Andrea Marra and Mara Floris	369
<i>Blaze-IT: A Lightweight BERT Model for the Italian Language</i> Francesco Russo and Michele Filannino	379
<i>Camoscio: An Italian Instruction-tuned LLaMA</i> Andrea Santilli and Emanuele Rodolà	385
<i>Automatic Generation of Common Procurement Vocabulary Codes</i> Lucia Siciliani, Emanuele Tanzi, Pierpaolo Basile and Pasquale Lops	396
<i>Inters8: A Corpus to Study Misogyny and Intersectionality on Twitter</i> Ivan Spada, Mirko Lai and Viviana Patti	403

<i>Let's Keep an Eye on Russian: Testing Sensitivity to the Change in the Grammatical Number in Somatic Idioms with ruBERT</i>	
Elena Spaziani and Eliana Di Palma	412
<i>“That branch of the Lake of Como...”: Developing a New Resource for the Analysis of I Promessi Sposi and its Historical Translations</i>	
Rachele Sprugnoli and Marco Sartor	420
<i>Qualitative Analysis of Persuasive Emotion Triggering in Online Content</i>	
Olga Uryupina	427
<i>When You Doubt, Abstain: A Study of Automated Fact-Checking in Italian under Domain Shift</i>	
Giovanni Valer, Alan Ramponi and Sara Tonelli	433
<i>On Incrementing Interpretability of Machine Learning Models from the Foundations: A Study on Syllabic Speech Units</i>	
Vincenzo Norman Vitale, Loredana Schettino and Francesco Cutugno	441
<i>Drug Name Recognition in the Cryptomarket Forum of Silk Road 2</i>	
Romane Werner, Thomas François and Sonja Bitzer	448
<i>Italian Crossword Generator: Enhancing Education through Interactive Word Puzzles</i>	
Kamyar Zeinalipour, Tommaso Iaquina, Asya Zanollo, Giovanni Angelini, Leonardo Rigutini, Marco Maggini and Marco Gori	455
<i>Die Rätselrevolution: Automated German Crossword Solving</i>	
Andrea Zugarini, Thomas Röthenbacher, Kai Klede, Marco Ernandes, Bjoern M. Eskofier and Dario Zanca	465
<i>Towards a Multilingual System for Vaccine Hesitancy Using a Data Mixture Approach</i>	
Oscar Araque, María Felipa Ledesma Corniel and Kyriaki Kalimeri	471
<i>TALL: A New Shiny App of Text Analysis for All</i>	
Massimo Aria, Corrado Cuccurullo, Luca D’Aniello, Michelangelo Misuraca and Maria Spano	476
<i>On the Impact of Language Adaptation for Large Language Models: A Case Study for the Italian Language Using Only Open Resources</i>	
Pierpaolo Basile, Pierluigi Cassotti, Marco Polignano, Lucia Siciliani and Giovanni Semeraro	480
<i>Annotating Homeric Emotions by a Domain-Specific Language</i>	
Federico Boschetti, Laura Chilla, Maria Konstantinidou and John Pavlopoulos	485
<i>How Good is NLLB for Low-resource Languages? A Study on the Genoese Language</i>	
Davide Buscaldi and Paolo Rosso	490
<i>Debunker Assistant: A Support for Detecting Online Misinformation</i>	
Arthur Thomas Edward Capozzi Lupi, Alessandra Teresa Cignarella, Simona Frenda, Mirko Lai, Marco Antonio Stranisci and Alessandra Urbinati	494
<i>Towards a Multi-Level Annotation Format for the Interoperability of Automatic Term Extraction Corpora</i>	
Nicola Cirillo and Daniela Vellutino	499
<i>The PBSDS: A Dataset for the Detection of Pseudoprofound Bullshit</i>	
Evan D. DeFrancesco and Carlo Strapparava	506

<i>A Post-Modern Approach to Automatic Metaphor Identification</i> Dario Del Fante, Federico Manzella, Guido Sciacvico and Eduard Ionel Stan	510
<i>Building a Corpus on Eating Disorders from TikTok: Challenges and Opportunities</i> Melissa Donati, Ludovica Polidori, Paola Vernillo and Gloria Gagliardi	515
<i>A Preliminary Release of the Italian Parliamentary Corpus</i> Valentino Frasnelli and Alessio Palmero Aprosio	521
<i>Extracting an Expectation-based Lexicon for UD Treebanks</i> Matteo Gay and Cristiano Chesi	525
<i>An Analysis of Visually Grounded Instructions in Embodied AI Tasks</i> Marco Grazioso and Alessandro Suglia	530
<i>Exploring Sentiments in Summarization: SentiTextRank, an Emotional Variant of TextRank</i> Md. Murad Hossain, Luca Anselma and Alessandro Mazzei	535
<i>An Italian Verb Lexicon for Sentiment Inference</i> Manfred Klenner, Dylan Massey and Susanna Tron	540
<i>The Inherence of Telicity: Unveiling Temporal Reasoning in Video Question Answering</i> Olga Loginova and Raffaella Bernardi	546
<i>Identification of Multiword Expressions: Comparing the Performance of a Conditional Random Fields Model on Corpora of Written and Spoken Italian</i> Ilaria Manfredi and Lorenzo Gregori	551
<i>Teasing LLMs Adapted to Italian</i> Leonardo Ranaldi, Giulia Pucci, Elena Sofia Ruzzetti, Fabio Massimo Zanzotto and André Freitas	557
<i>Investigating Gender Bias in Large Language Models for the Italian Language</i> Elena Sofia Ruzzetti, Dario Onorati, Leonardo Ranaldi, Davide Venditti and Fabio Massimo Zanzotto	562
<i>Towards a New Computational Lexicon for Italian: Building the Morphological Layer by Harmonizing and Merging Existing Resources</i> Flavia Sciolette, Simone Marchi and Emiliano Giovannetti	570
<i>Textual Entailment with Natural Language Explanations: The Italian e-RTE-3 Dataset</i> Andrea Zaninello, Sofia Brenna and Bernardo Magnini	575