

CLiC-it 2020

The Seventh Italian Conference on Computational Linguistics

Proceedings of the Conference

March 1-3, 2021

Copyright ©2020 for the individual papers by the papers' authors.
Copyright ©2020 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-2769>.
The papers are mirrored in the ACL Anthology.

ISBN 979-12-80136-28-2

Table of Contents

<i>Preface</i>	
Johanna Monti, Felice Dell’Orletta and Fabio Tamburini	1
<i>Fine-grained sentiment analysis: a piece of cake?</i>	
Veronique Hoste	6
<i>Interaction-aware multimodal dialogue with conversational agents</i>	
Stefan Kopp	8
<i>Distributional Semantics: Yesterday, Today, and Tomorrow</i>	
Alessandro Lenci	9
<i>Investigating Proactivity in Task-Oriented Dialogues</i>	
Vevake Balaraman and Bernardo Magnini	10
<i>Domain Adaptation for Text Classification with Weird Embeddings</i>	
Valerio Basile	18
<i>Personal-ITY: A Novel YouTube-based Corpus for Personality Prediction in Italian</i>	
Elisa Bassignana, Malvina Nissim and Viviana Patti	25
<i>Gender Bias in Italian Word Embeddings</i>	
Davide Biasion, Alessandro Fabris, Gianmaria Silvello and Gian Antonio Susto	32
<i>Automatic Induction of FrameNet lexical units in Italian</i>	
Silvia Brambilla, Danilo Croce, Fabio Tamburini and Roberto Basili	40
<i>Detecting Early Onset of Depression from Social Media Text using Learned Confidence Scores</i>	
Ana-Maria Bucur and Liviu Dinu	47
<i>Building a Treebank in Universal Dependencies for Italian Sign Language</i>	
Gaia Caligiore, Cristina Bosco and Alessandro Mazzei	53
<i>Hate Speech Detection with Machine-Translated Data: The Role of Annotation Scheme, Class Imbalance and Undersampling</i>	
Camilla Casula and Sara Tonelli	59
<i>UDante: First Steps Towards the Universal Dependencies Treebank of Dante’s Latin Works</i>	
Flavio Massimiliano Cecchini, Rachele Sprugnoli, Giovanni Moretti and Marco Passarotti	66
<i>Spotto la quarantena”: per una analisi dell’italiano scritto degli studenti universitari via social network in tempo di COVID-19</i>	
Francesca Chiusaroli, Johanna Monti, Maria Laura Pierucci and Gennaro Nolano	73
<i>Italian Counter Narrative Generation to Fight Online Hate Speech</i>	
Yi-Ling Chung, Serra Sinem Tekirođlu and Marco Guerini	82
<i>Clustering verbal Objects: Manual and Automatic Procedures Compared</i>	
Ilaria Colucci, Elisabetta Jeřek and Vít Baisa	89
<i>GePpeTto Carves Italian into a Language Model</i>	
Lorenzo De Mattei, Michele Cafagna, Felice Dell’Orletta, Malvina Nissim and Marco Guerini	97
<i>Phonological Layers of Meaning: A Computational Exploration of Sound Iconicity</i>	
Andrea Gregor de Varda and Carlo Strapparava	105

<i>Natural Language Generation in Dialogue Systems for Customer Care</i>	
Mirko Di Lascio, Manuela Sanguinetti, Luca Anselma, Dario Mana, Alessandro Mazzei, Viviana Patti and Rossana Simeoni	112
<i>Cross-Language Transformer Adaptation for Frequently Asked Questions</i>	
Luca Di Liello, Daniele Bonadiman, Alessandro Moschitti, Cristina Giannone, Andrea Favalli and Raniero Romagnoli	118
<i>How Good are Humans at Native Language Identification? A Case Study on Italian L2 writings</i>	
Elisa Di Nuovo, Cristina Bosco and Elisa Corino	125
<i>Linguistic Features in Automatic Sarcasm Detection</i>	
Martina Ducret, Lauren Kruse, Carlos Martinez, Anna Feldman and Jing Peng	132
<i>Risorse e strumenti per le varietà storiche dell’italiano: il progetto TrAVaSI</i>	
Manuel Favaro, Marco Biffi and Simonetta Montemagni	139
<i>The AEREST Reading Database</i>	
Marcello Ferro, Sara Giulivi and Claudia Cappa	148
<i>On Knowledge Distillation for Direct Speech Translation</i>	
Marco Gaido, Mattia Antonino Di Gangi, Matteo Negri and Marco Turchi	154
<i>Grounded and Ungrounded Referring Expressions in Human Dialogues: Language Mirrors Different Grounding Conditions</i>	
Eleonora Gualdoni, Raffaella Bernardi, Raquel Fernández and Sandro Pezzelle	161
<i>Point Break: Surfing Heterogeneous Data for Subtitle Segmentation</i>	
Alina Karakanta, Matteo Negri and Marco Turchi	168
<i>How Granularity of Orthography-Phonology Mappings Affect Reading Development: Evidence from a Computational Model of English Word Reading and Spelling</i>	
Alfred Lim, Beth A. O’Brien and Luca Onnis	175
<i>Simple Data Augmentation for Multilingual NLU in Task Oriented Dialogue Systems</i>	
Samuel Louvan and Bernardo Magnini	182
<i>The E3C Project: Collection and Annotation of a Multilingual Corpus of Clinical Cases</i>	
Bernardo Magnini, Begoña Altuna, Alberto Lavelli, Manuela Speranza and Roberto Zanolini . .	190
<i>Monitoring Social Media to Identify Environmental Crimes through NLP. A preliminary study</i>	
Raffaele Manna, Antonio Pascucci, Wanda Punzi Zarino, Vincenzo Simoniello and Johanna Monti	197
<i>Does Finger-Tracking Point to Child Reading Strategies?</i>	
Claudia Marzi, Anna Rodella, Andrea Nadalini, Loukia Taxitari and Vito Pirrelli	204
<i>The Style of a Successful Story: a Computational Study on the Fanfiction Genre</i>	
Andrea Mattei, Dominique Brunato and Felice Dell’Orletta	211
<i>A Multimodal Dataset of Images and Text to Study Abusive Language</i>	
Stefano Menini, Alessio Palmero Aprosio and Sara Tonelli	217
<i>A Resource for Detecting Misspellings and Denoising Medical Text Data</i>	
Enrico Mensa, Gian Manuel Marino, Davide Colla, Matteo Delsanto and Daniele P. Radicioni	223
<i>Is Neural Language Model Perplexity Related to Readability?</i>	
Alessio Miaschi, Chiara Alzetta, Dominique Brunato, Felice Dell’Orletta and Giulia Venturi	230

<i>Italian Transformers Under the Linguistic Lens</i>	
Alessio Miaschi, Gabriele Sarti, Dominique Brunato, Felice Dell’Orletta and Giulia Venturi .	237
<i>ItaGLAM: A corpus of Cultural Communication on Twitter during the Pandemic</i>	
Gennaro Nolano, Carola Carlino, Maria Pia Di Buono and Johanna Monti	244
<i>(Stem and Word) Predictability in Italian Verb Paradigms: An Entropy-Based Study Exploiting the New Resource LeFFI</i>	
Matteo Pellegrini and Alessandra Teresa Cignarella	251
<i>A Case Study of Natural Gender Phenomena in Translation. A Comparison of Google Translate, Bing Microsoft Translator and DeepL for English to Italian, French and Spanish</i>	
Argentina Anna Rescigno, Eva Vanmassenhove, Johanna Monti and Andy Way	257
<i>Multifunctional ISO Standard Dialogue Act Tagging in Italian</i>	
Gabriel Roccabruna, Alessandra Cervone and Giuseppe Riccardi	263
<i>Tracing Metonymic Relations in T-PAS: An Annotation Exercise on a Corpus-based Resource for Italian</i>	
Emma Romani and Elisabetta Ježek	271
<i>Datasets and Models for Authorship Attribution on Italian Personal Writings</i>	
Gaetana Ruggiero, Albert Gatt and Malvina Nissim	278
<i>The Archaeo-Term Project: Multilingual Terminology in Archaeology</i>	
Giulia Speranza, Raffaele Manna, Maria Pia Di Buono and Johanna Monti	285
<i>Exploiting Distributional Semantics Models for Natural Language Context-aware Justifications for Recommender Systems</i>	
Giuseppe Spillo, Cataldo Musto, Marco de Gemmis, Pasquale Lops and Giovanni Semeraro .	292
<i>MultiEmotions-It: a New Dataset for Opinion Polarity and Emotion Analysis for Italian</i>	
Rachele Sprugnoli	300
<i>Becoming JILDA</i>	
Irene Sucameli, Alessandro Lenci, Bernardo Magnini, Maria Simi and Manuela Speranza . . .	307
<i>How BERTology Changed the State-of-the-Art also for Italian NLP</i>	
Fabio Tamburini	313
<i>Overprotective Training Environments Fall Short at Testing Time: Let Models Contribute to Their Own Training</i>	
Alberto Testoni and Raffaella Bernardi	320
<i>Topic Modelling Games</i>	
Rocco Tripodi	326
<i>Polarity Imbalance in Lexicon-based Sentiment Analysis</i>	
Marco Vassallo, Giuliano Gabrieli, Valerio Basile and Cristina Bosco	334
<i>#andràtuttobene: Images, Texts, Emojis and Geodata in a Sentiment Analysis Pipeline</i>	
Pierluigi Vitale, Serena Pelosi and Mariacristina Falco	341
<i>Suoidne-varra-bleahkka-mála-bihkka-senet-dielku ’hay-blood-ink-paint-tar-mustard-stain’ -Should compounds be lexicalized in NLP?</i>	
Linda Wiechetek, Chiara Argese, Tommi A Pirinen and Trond Trosterud	348

<i>Quantitative Linguistic Investigations across Universal Dependencies Treebanks</i>	
Chiara Alzetta, Felice Dell’Orletta, Simonetta Montemagni, Petya Osenova, Kiril Simov and Giulia Venturi	355
<i>A Machine Learning approach for Sentiment Analysis for Italian Reviews in Healthcare</i>	
Luca Bacco, Andrea Cimino, Luca Paulon, Mario Merone and Felice Dell’Orletta	362
<i>A Diachronic Italian Corpus based on L’Unità"</i>	
Pierpaolo Basile, Annalina Caputo, Tommaso Caselli, Pierluigi Cassotti and Rossella Varvara	369
<i>The Corpus Anchise 320 and the Analysis of Conversations between Healthcare Workers and People with Dementia</i>	
Nicola Benvenuti, Andrea Bolioli, Alessandro Mazzei, Pietro Vigorelli and Alessio Bosca...	375
<i>Analysis of Lexical Semantic Changes in Corpora with the Diachronic Engine</i>	
Pierluigi Cassotti, Pierpaolo Basile, Marco de Gemmis and Giovanni Semeraro	382
<i>Surviving the Legal Jungle: Text Classification of Italian Laws in Extremely Noisy Conditions</i>	
Riccardo Coltrinari, Alessandro Antinori and Fabio Celli	389
<i>AriEmozione: Identifying Emotions in Opera Verses</i>	
Francesco Fericola, Shibingfeng Zhang, Federico Garcea, Paolo Bonora and Alberto Barrón-Cedeño	395
<i>Gracissare: Ancient Greek Loanwords in the LiLa Knowledge Base of Linguistic Resources for Latin</i>	
Greta Franzini, Federica Zampedri, Marco Passarotti, Francesco Mambrini and Giovanni Moretti	401
<i>L’impatto emotivo della comunicazione istituzionale durante la pandemia di COVID-19: uno studio di Twitter Sentiment Analysis</i>	
Gloria Gagliardi, Lorenzo Gregori and Alice Suozzi	407
<i>Predicting Social Exclusion: A Study of Linguistic Ostracism in Social Networks</i>	
Greta Gandolfi and Carlo Strapparava	413
<i>Predicting Movie-elicited Emotions from Dialogue in Screenplay Text: A Study on "Forrest Gump"</i>	
Benedetta Iavarone and Felice Dell’Orletta	418
<i>Multiword Expressions We Live by: A Validated Usage-based Dataset from Corpora of Written Italian</i>	
Francesca Masini, M. Silvia Micheli, Andrea Zaninello, Sara Castagnoli and Malvina Nissim	424
<i>BERTino: An Italian DistilBERT model</i>	
Matteo Muffo and Enrico Bertino	429
<i>Creativity Embedding: A Vector to Characterise and Classify Plausible Triples in Deep Learning NLP Models</i>	
Isabeau Oliveri, Luca Ardito, Giuseppe Rizzo and Maurizio Morisio	435
<i>The CREENDER Tool for Creating Multimodal Datasets of Images and Comments</i>	
Alessio Palmero Aprosio, Stefano Menini and Sara Tonelli	441
<i>A Deep Learning Model for the Analysis of Medical Reports in ICD-10 Clinical Coding Task</i>	
Marco Polignano, Pierpaolo Basile, Marco de Gemmis, Pasquale Lops and Giovanni Semeraro	446
<i>Exploring Attention in a Multimodal Corpus of Guided Tours</i>	
Andrea Amelio Ravelli, Antonio Origlia and Felice Dell’Orletta	452

<i>Valutazione umana di DeepL a livello di frase per le traduzioni di testi specialistici dall'inglese verso l'italiano</i>	
Mirko Tivosanis and Sirio Papa	458
<i>Dialog-based Help Desk through Automated Question Answering and Intent Detection</i>	
Antonio Uva, Pierluigi Roberti and Alessandro Moschitti	465
<i>Analyses of Character Emotions in Dramatic Works by Using EmoLex Unigrams</i>	
Mehmet Can Yavuz	472