

CLiC-it 2015

The Second Italian Conference on Computational Linguistics

Proceedings of the Conference

December 3-4, 2015

Copyright ©2015 by AILC – Associazione Italiana di Linguistica Computazionale.
The text of this volume is available under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0). Other elements may be subject to separate rights, as stated by OpenEdition.

These proceedings were published by Accademia University Press.
A digital edition is available at:

<https://books.openedition.org/aaccademia/1277>

The papers are mirrored in the ACL Anthology.

ISBN 978-88-99200-62-6

Table of Contents

Preface

Cristina Bosco, Sara Tonelli and Fabio Massimo Zanzotto

Bolzano/Bozen Corpus: Coding Information about the Speaker in IMDI Metadata Structure

Marco Angster 9

Detecting the scope of negations in clinical notes

Giuseppe Attardi, Vittoria Cozza and Daniele Sartiano 14

Deep Learning for Social Sensing from Tweets

Giuseppe Attardi, Laura Gorrieri, Alessio Miaschi and Ruggero Petrolito 20

Evolution of Italian Treebank and Dependency Parsing towards Universal Dependencies

Giuseppe Attardi, Simone Saletti and Maria Simi 25

CIa: un Corpus di Produzioni Scritte di Apprendenti l'Italiano L1 Annotato con Errori

A. Barbagli, P. Lucisano, F. Dell'Orletta, S. Montemagni and G. Venturi 31

Entity Linking for Italian Tweets

Pierpaolo Basile, Annalina Caputo and Giovanni Semeraro 36

Deep Tweets: from Entity Linking to Sentiment Analysis

Pierpaolo Basile, Valerio Basile, Malvina Nissim and Nicole Novielli 41

Enhancing the Accuracy of Ancient Greek WordNet by Multilingual Distributional Semantics

Yuri Bizzoni, Riccardo Del Gratta, Federico Boschetti and Marianne Reboul 47

Deep Neural Networks for Named Entity Recognition in Italian

Daniele Bonadiman, Aliaksei Severyn and Alessandro Moschitti 51

Exploring Cross-Lingual Sense Mapping in a Multilingual Parallel Corpus

Francis Bond and Giulia Bonansinga 56

ISACCO: a corpus for investigating spoken and written language development in Italian school-age children

Dominique Brunato and Felice Dell'Orletta 62

Inconsistencies Detection in Bipolar Entailment Graphs

Elena Cabrio and Serena Villata 67

A Graph-based Model of Contextual Information in Sentiment Analysis over Twitter

Giuseppe Castellucci, Danilo Croce and Roberto Basili 72

Word Sense Discrimination: A gangplank algorithm

Flavio Massimiliano Cecchini and Elisabetta Fersini 77

Facebook and the Real World: Correlations between Online and Offline Conversations

Fabio Celli and Luca Polonio 82

La scrittura in emoji tra dizionario e traduzione

Francesca Chiusaroli 88

<i>On Mining Citations to Primary and Secondary Sources in Historiography</i> Giovanni Colavizza and Frédéric Kaplan	94
<i>Visualising Italian Language Resources: a Snapshot</i> Riccardo Del Gratta, Francesca Frontini, Monica Monachini, Gabriella Pardelli, Irene Russo, Roberto Bartolini, Sara Goggi, Fahad Khan, Valeria Quochi, Claudia Soria and Nicoletta Calzolari	100
<i>A manually-annotated Italian corpus for fine-grained sentiment analysis</i> Marilena Di Bari, Serge Sharoff and Martin Thomas	105
<i>From a Lexical to a Semantic Distributional Hypothesis</i> Luigi Di Caro, Guido Boella, Alice Ruggeri, Loredana Cupi, John Adebayo Kolawole and Livio Robaldo	110
<i>An Active Learning Approach to the Classification of Non-Sentential Utterances</i> Paolo Dragone and Pierre Lison	115
<i>The CompWHoB Corpus: Computational Construction, Annotation and Linguistic Analysis of the White House Press Briefings Corpus</i> Fabrizio Esposito, Pierpaolo Basile, Francesco Cutugno and Marco Venuti	120
<i>Costituzione di un corpus giuridico parallelo italiano-arabo</i> Fathi Fawi	125
<i>Italian-Arabic domain terminology extraction from parallel corpora</i> Fathi Fawi and Rodolfo Delmonte	130
<i>Annotating opposition among verb senses: a crowdsourcing experiment</i> Anna Feltracco, Elisabetta Jezek, Bernardo Magnini and Simone Magnolini	135
<i>Gold standard vs. silver standard: the case of dependency parsing for Italian</i> Michele Filannino and Marilena Di Bari	141
<i>Phrase Structure and Ancient Anatolian languages. Methodology and challenges for a Luwian syntactic annotation</i> Federico Giusfredi	146
<i>Linking dei contenuti multimediali tra ontologie multilingui: i verbi di azione tra IMAGACT e BabelNet</i> Lorenzo Gregori, Andrea Amelio Ravelli and Alessandro Panunzi	150
<i>New wine in old wineskins: a morphology-based approach to translate medical terminology</i> Raffaele Guarasci and Alessandro Maisto	155
<i>Computing, memory and writing: some reflections on an early experiment in digital literary studies</i> Giorgio Guzzetta and Federico Nanni	161
<i>Effectiveness of Domain Adaptation Approaches for Social Media PoS Tagging</i> Tobias Horsmann and Torsten Zesch	166
<i>Building a Corpus on a Debate on Political Reform in Twitter</i> Mirko Lai, Daniela Virone, Cristina Bosco and Viviana Patti	171
<i>The OPATCH corpus platform – facing heterogeneous groups of texts and users</i> Verena Lyding, Michel Génèreux, Katalin Szabò and Johannes Andresen	177

<i>Generare messaggi persuasivi per una dieta salutare</i> Alessandro Mazzei	182
<i>FacTA: Evaluation of Event Factuality and Temporal Anchoring</i> Anne-Lyse Minard, Manuela Speranza, Rachele Sprugnoli and Tommaso Caselli ...	187
<i>TED-MWE: a bilingual parallel corpus with MWE annotation. Towards a methodology for annotating MWEs in parallel multilingual corpora</i> Johanna Monti, Federico Sangati and Mihael Arcan	193
<i>Digging in the Dirt: Extracting Keyphrases from Texts with KD</i> Giovanni Moretti, Rachele Sprugnoli and Sara Tonelli	198
<i>Automatic extraction of Word Combinations from corpora: evaluating methods and benchmarks</i> Malvina Nissim, Sara Castagnoli, Francesca Masini, Gianluca E. Lebani, Lucia Passaro and Alessandro Lenci	204
<i>Improved Written Arabic Word Parsing through Orthographic, Syntactic and Semantic constraints</i> Ouafae Nahli and Simone Marchi	210
<i>ItEM: A Vector Space Model to Bootstrap an Italian Emotive Lexicon</i> Lucia C. Passaro, Laura Pollacci and Alessandro Lenci	215
<i>Somewhere between Valency Frames and Synsets. Comparing Latin Vallex and Latin WordNet</i> Marco Passarotti, Berta González Saavedra and Christophe Onambélé Manga	221
<i>SentIta and Doxa: Italian Databases and Tools for Sentiment Analysis Purposes</i> Serena Pelosi	226
<i>Le scritture brevi dello storytelling: analisi di case studies di successo</i> Maria Laura Pierucci	232
<i>Tracking the Evolution of Written Language Competence: an NLP-based Approach</i> Stefan Richter, Andrea Cimino, Felice Dell’Orletta and Giulia Venturi	236
<i>Learning Grasping Possibilities for Artifacts: Dimensions, Weights and Distributional Semantics</i> Irene Russo and Irene De Felice	241
<i>Experimenting the use of catenae in Phrase-Based SMT</i> Manuela Sanguinetti	246
<i>Cross-language projection of multilayer semantic annotation in the NewsReader Wikinews Italian Corpus (WItaC)</i> Manuela Speranza and Anne-Lyse Minard	252
<i>Parsing Events: a New Perspective on Old Challenges</i> Rachele Sprugnoli, Felice Dell’Orletta, Tommaso Caselli, Simonetta Montemagni and Cristina Bosco	258
<i>Generalization in Native Language Identification: Learners versus Scientists</i> Sabrina Stehwien and Sebastian Padó	264
<i>Sentiment Polarity Classification with Low-level Discourse-based Features</i> Evgeny A. Stepanov and Giuseppe Riccardi	269

<i>Analyzing and annotating for sentiment analysis the socio-political debate on #labuonascuola</i>	
Marco Stranisci, Cristina Bosco, Viviana Patti and Delia Irazú Hernández Farías . . .	274
<i>Reference-free and Confidence-independent Binary Quality Estimation for Automatic Speech Recognition</i>	
Hamed Zamani, José G. C. de Souza, Matteo Negri, Marco Turchi and Daniele Falavigna	280