

CLiC-it 2016

The Third Italian Conference on Computational Linguistics

Proceedings of the Conference

December 5-6, 2016

Copyright ©2016 for the individual papers by the papers' authors.
Copyright ©2016 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-1749>.
The papers are mirrored in the ACL Anthology.

ISBN 978-88-99982-08-9

Table of Contents

<i>Preface</i>	
Anna Corazza, Simonetta Montemagni and Giovanni Semeraro	1
<i>Reflections on Universal Dependencies</i>	
Joakim Nivre	10
<i>What's this movie about? Automatic Content Analysis and Summarization</i>	
Mirella Lapata	11
<i>Determining the Compositionality of Noun-Adjective Pairs with Lexical Variants and Distributional Semantics</i>	
Marco S. G. Senaldi, Gianluca E. Leboni and Alessandro Lenci	12
<i>Studio sull'Ordine dei Costituenti nel Confronto tra Generi e Complessità (Analysis of Constituents Order Across Textual Genres and Complexity)</i>	
Giulia Pieri, Dominique Brunato and Felice Dell'Orletta	18
<i>Grounding the Lexical Sets of Causative-Inchoative Verbs with Word Embedding</i>	
Edoardo Maria Ponti, Elisabetta Jezek and Bernardo Magnini	23
<i>Dieci sfumature di marcatezza sintattica: verso una nozione computazionale di complessità (Ten Shades of Syntactic Markedness: Towards a Computational Notion of Complexity)</i>	
Erica Tusa, Felice Dell'Orletta, Simonetta Montemagni and Giulia Venturi	29
<i>Online Automatic Post-Editing across Domains</i>	
Rajen Chatterjee, Gebremedhen Gebremelak, Matteo Negri and Marco Turchi	35
<i>Language Resources for Italian: towards the Development of a Corpus of Annotated Italian Multiword Expressions</i>	
Shiva Taslimipoor, Anna Desantis, Manuela Cherchi, Ruslan Mitkov and Johanna Monti	41
<i>Emojitalianobot and EmojiWorldBot - New online Tools and Digital Environments for Translation into Emoji</i>	
Johanna Monti, Federico Sangati, Francesca Chiusaroli, Martin Benjamin and Sina Mansour	47
<i>IR Scientific Data: How to Semantically Represent and Enrich Them</i>	
Toine Bogers, Georgeta Bordea, Paul Buitelaar, Nicola Ferro and Gianmaria Silvello	52
<i>Nyström Methods for Efficient Kernel-Based Methods for Community Question Answering</i>	
Danilo Croce, Simone Filice and Roberto Basili	58
<i>Topic Modelling with Word Embeddings</i>	
Fabrizio Esposito, Anna Corazza and Francesco Cutugno	64
<i>Imparare a Quantificare Guardando (Learning to Quantify by Watching)</i>	
Sandro Pezzelle, Ionut Sorodoc, Aurelie Herbelot and Raffaella Bernardi	70
<i>Appetitoso: A Search Engine for Restaurant Retrieval based on Dishes</i>	
Gianni Barlacchi, Azad Abad, Emanuele Rossinelli and Alessandro Moschitti	76
<i>Gamification for IR: The Query Aspects Game</i>	
Giorgio Maria Di Nunzio, Maria Maistro and Daniel Zilio	81

<i>Semantic Interpretation of Events in Live Soccer Commentaries</i>	
Anne-Lyse Minard, Manuela Speranza, Bernardo Magnini and Mohammed R. H. Qwaider . . .	87
<i>Linking IMAGACT Ontology to BabelNet through Action Videos</i>	
Lorenzo Gregori, Alessandro Panunzi and Andrea Amelio Ravelli	93
<i>Relation Mining from Clinical Records</i>	
Anita Alicante, Anna Corazza, Francesco Isgrò and Stefano Silvestri	99
<i>Subjective Well-Being and Social Media: A Semantically Annotated Twitter Corpus on Fertility and Parenthood</i>	
Emilio Sulis, Cristina Bosco, Viviana Patti, Mirko Lai, Delia Irazú Hernández Farías, Letizia Mencarini, Michele Mozzachiodi and Daniele Vignoli	104
<i>An Extended Version of the KoKo German L1 Learner Corpus</i>	
Andrea Abel, Aivars Glaznieks, Lionel Nicolas and Egon W. Stemle	110
<i>Annotating Content Zones in News Articles</i>	
Daniela Baiamonte, Tommaso Caselli and Irina Prodanof	116
<i>LICO: A Lexicon of Italian Connectives</i>	
Anna Feltracco, Elisabetta Jezek, Bernardo Magnini and Manfred Stede	122
<i>Formatio formosa est. Building a Word Formation Lexicon for Latin</i>	
Eleonora Litta, Marco Passarotti and Chris Culy	127
<i>SIMPITIKI: a Simplification corpus for Italian</i>	
Sara Tonelli, Alessio Palmero Aprosio and Francesca Saltori	132
<i>(Almost) Automatic Conversion of the Venice Italian Treebank into the Merged Italian Dependency Treebank Format</i>	
Linda Alfieri and Fabio Tamburini	138
<i>Enriching the Ita-TimeBank with Narrative Containers</i>	
Alice Bracchi, Tommaso Caselli and Irina Prodanof	143
<i>Increasing Information Accessibility on the Web: a Rating System for Specialized Dictionaries</i>	
Valeria Caruso, Anna De Meo and Vincenzo Norman Vitale	149
<i>Parsing di Corpora di Apprendenti di Italiano: un Primo Studio su VALICO (Parsing Italian Learner Corpora: a Case Study on VALICO)</i>	
Elisa Corino and Claudio Russo	154
<i>Spammare senza pietà - Corpus based analysis of English, unacclimatised verb Loans in Italian and Creation of a Reference Lexicon</i>	
Anna Fantini	160
<i>Taming Sense Sparsity: a Common-Sense Approach</i>	
Antonio Lieto, Enrico Mensa and Daniele P. Radicioni	166
<i>Building a Computational Lexicon by using SQL</i>	
Alessandro Mazzei	172
<i>FB-NEWS15: A Topic-Annotated Facebook Corpus for Emotion Detection and Sentiment Analysis</i>	
Lucia C. Passaro, Alessandro Bondielli and Alessandro Lenci	177
<i>Sardinian on Facebook: Analysing Diatopic Varieties through Translated Lexical Lists</i>	
Irene Russo, Simone Pisano and Claudia Soria	182

<i>A New Dataset for Source Code Comment Coherence</i>	
Anna Corazza, Valerio Maggio and Giuseppe Scanniello	187
<i>Tracing Metaphors in Time through Self-Distance in Vector Spaces</i>	
Marco Del Tredici, Malvina Nissim and Andrea Zaninello	192
<i>KD Strikes Back: from Keyphrases to Labelled Domains Using External Knowledge Sources</i>	
Giovanni Moretti, Rachele Sprugnoli and Sara Tonelli	198
<i>Investigating the Application of Distributional Semantics to Stylometry</i>	
Giulia Benotto, Emiliano Giovannetti and Simone Marchi	204
<i>Diachronic Analysis of the Italian Language exploiting Google Ngram</i>	
Pierpaolo Basile, Annalina Caputo, Roberta Luisi and Giovanni Semeraro	209
<i>Hybrid Language Segmentation for Historical Documents</i>	
David Alfter and Yuri Bizzoni	214
<i>La Leggibilità dei Testi di Ambito Medico rivolti al Paziente: il Caso dei Bugiardini di Farmaci Senza Obbligo di Prescrizione Medica (Readability of Medical Texts for Patients: a Case Study on Information Leaflets of Drugs without Obligation of Medical Prescription)</i>	
Franca Orletti, Felice Dell’Orletta and Rossella Iovino	219
<i>La Modellazione Diacronica di Risorse Termino-Ontologiche nell’Ambito delle Digital Humanities: Esperimenti su Clavius (Diachronic Modeling of Terminological-Ontological Resources within Digital Humanities: Experiments on Clavius)</i>	
Silvia Piccini, Andrea Bellandi, Giulia Benotto and Emiliano Giovannetti	225
<i>Panta Rei: Tracking Semantic Change with Distributional Semantics in Ancient Greek</i>	
Martina A. Rodda, Marco S. G. Senaldi and Alessandro Lenci	230
<i>Reassessing Inflectional Regularity in Modern Greek Conjugation</i>	
Stavros Bompolas, Marcello Ferro, Claudia Marzi, Franco Alberto Cardillo and Vito Pirrelli	235
<i>Written Word Production and Lexical Self-Organisation: Evidence from English (Pseudo)Compounds</i>	
Marcello Ferro, Franco Alberto Cardillo, Vito Pirrelli, Christina L. Gagné and Thomas L. Spalding	241
<i>Lexical Categories or Frequency Effects? A Feedback from Quantitative Methods Applied to Psycholinguistic Models in two Studies on Italian</i>	
Francesca Franzon, Giorgio Arcara and Chiara Zanini	247
<i>Semantic Priming Effects in Italian Verbs Recognition: the Role of Grammatical Classes and Semantic Categories</i>	
Azzurra Mancuso, Maria De Martino and Alessandro Laudanna	252
<i>Sentiment Analysis: Applicazione in un Dominio Psico-Forense</i>	
Francesca Guglielmi, Pierpaolo Basile, Antonietta Curci and Giovanni Semeraro	257
<i>Stepping out of the Chinese Room: Word meaning with and without consciousness</i>	
Roberto Bottini, Daniel Casasanto, Andrea Nadalini and Davide Crepaldi	262
<i>Context-aware Spoken Language Understanding for Human Robot Interaction</i>	
Andrea Vanzo, Danilo Croce, Roberto Basili and Daniele Nardi	267
<i>Argument Mining on Italian News Blogs</i>	
Pierpaolo Basile, Valerio Basile, Elena Cabrio and Serena Villata	273

<i>Twitter Sentiment Polarity Classification using Barrier Features</i>	
Anita Alicante, Anna Corazza and Antonio Pironti	278
<i>The DiDi Corpus of South Tyrolean CMC Data: A multilingual corpus of Facebook texts</i>	
Jennifer-Carmen Frey, Aivars Glaznieks and Egon W. Stemle	284
<i>Sequenze N+pN (Nome Comune + Nome Proprio): Descrizione Linguistica da un Corpus dell’Italiano</i>	
<i>(N+pN (Common Noun + Proper Noun) Sequences: Linguistic Description from an Italian Corpus)</i>	
Felicia Logozzo	289
<i>(Better than) State-of-the-Art PoS-tagging for Italian Texts</i>	
Fabio Tamburini	294
<i>Tree Kernels-based Discriminative Reranker for Italian Constituency Parsers</i>	
Antonio Uva and Alessandro Moschitti	299
<i>Comparing State-of-the-art Dependency Parsers on the Italian Stanford Dependency Treebank</i>	
Alberto Lavelli	304
<i>May the Goddess of Hope Help Us. Homonymy in Latin Lexicon and Onomasticon</i>	
Marco Passarotti and Marco Budassi	310