

CLiC-it 2014

The First Italian Conference on Computational Linguistics

Proceedings of the Conference

December 9-10, 2014

Copyright ©2014 by Pisa University Press srl.

These proceedings were published by Pisa University Press.

The original papers are available at:

<https://clic2014.fileli.unipi.it/it/atti>

The papers are mirrored in the ACL Anthology.

ISBN 978-886741-472-7

Table of Contents

Preface

Roberto Basili, Alessandro Lenci and Bernardo Magnini

Creating a standard for evaluating distant supervision for relation extraction

Azad Abad and Alessandro Moschitti 1

Towards Compositional Tree Kernels

Paolo Annesi, Danilo Croce and Roberto Basili 7

Initial explorations in Kazakh to English statistical machine translation

Zhenisbek Assylbekov and Assulan Nurkas 12

Adapting linguistic tools for the analysis of Italian medical records

Giuseppe Attardi, Vittoria Cozza and Daniele Sartiano 17

Tecnologie del linguaggio e monitoraggio dell'evoluzione delle abilità di scrittura nella scuola secondaria di primo grado

Alessia Barbagli, Pietro Lucisano, Felice Dell'Orletta, Simonetta Montemagni and Giulia Venturi
23

Italian irony detection in Twitter: A first approach

Francesco Barbieri, Francesco Ronzano and Horacio Saggion 28

A retrieval model for automatic resolution of crossword puzzles in Italian language

Gianni Barlacchi, Massimo Nicosia and Alessandro Moschitti 33

Analysing word meaning over time by exploiting temporal Random Indexing

Pierpaolo Basile, Annalina Caputo and Giovanni Semeraro 38

Combining distributional semantic models and sense distribution for effective Italian word sense disambiguation

Pierpaolo Basile, Annalina Caputo and Giovanni Semeraro 43

A Lesk-inspired unsupervised algorithm for lexical choice from WordNet synsets

Valerio Basile 48

The Talmud System: a Collaborative web Application for the Translation of the Babylonian Talmud Into Italian

Andrea Bellandi, Davide Albanesi, Alessia Bellusci, Andrea Bozzi and Emiliano Giovannetti . 53

Towards a Decision Support System for Text Interpretation

Alessia Bellusci, Andrea Bellandi, Giulia Benotto, Amedeo Cappelli, Emiliano Giovannetti and Simone Marchi 58

An Italian dataset of textual entailment graphs for text exploration of customer interactions

Luisa Bentivogli and Bernardo Magnini 63

L'integrazione di informazioni contestuali e linguistiche nel riconoscimento automatico dell'ironia

Lorenzo Bernardini and Irina Prodanof 67

A generic tool for the automatic syllabification of Italian

Brigitte Bigi and Caterina Petrone 73

<i>Errori di OCR e riconoscimento di entità nell'Archivio Storico de La Stampa</i>	
Andrea Bolioli, Eleonora Marchioni and Raffaella Ventaglio	78
<i>Computer assisted annotation of themes and motifs in ancient Greek epigrams: First steps</i>	
Federico Boschetti, Riccardo Del Gratta and Marion Lamé	83
<i>Defining an annotation scheme with a view to automatic text simplification</i>	
Dominique Brunato, Felice Dell'Orletta, Giulia Venturi and Simonetta Montemagni	87
<i>Senso Comune as a knowledge base of Italian language: the resource and its development</i>	
Tommaso Caselli, Isabella Chiari, Aldo Gangemi, Elisabetta Jezek, Alessandro Oltramari, Guido Vetere, Laure Vieu and Fabio Massimo Zanzotto	93
<i>CorEA: Italian news corpus with emotions and agreement</i>	
Fabio Celli, Giuseppe Riccardi and Arindam Ghosh	98
<i>Detecting attribution relations in speech: A corpus study</i>	
Alessandra Cervone, Silvia Pareti, Peter Bell, Irina Prodanof and Tommaso Caselli	103
<i>Adattamento al progetto dei modelli di traduzione automatica nella traduzione assistita</i>	
Mauro Cettolo, Nicola Bertoldi and Marcello Federico	108
<i>The new basic vocabulary of Italian as a linguistic resource</i>	
Isabella Chiari and Tullio De Mauro	113
<i>Sintassi e semantica dell'hashtag: studio preliminare di una forma di Scritture Brevi</i>	
Francesca Chiusaroli	117
<i>Annotation of complex emotions in real-life dialogues: The case of empathy</i>	
Morena Danieli, Giuseppe Riccardi and Firoj Alam	122
<i>Evaluating ImagAct-WordNet mapping for English and Italian through videos</i>	
Irene De Felice, Roberto Bartolini, Irene Russo, Valeria Quochi and Monica Monachini	128
<i>CLaSSES: a new digital resource for Latin epigraphy</i>	
Irene De Felice, Margherita Donati and Giovanna Marotta	132
<i>Online and Multitask Learning for Machine Translation Quality Estimation in Real-world Scenarios</i>	
José G. C. de Souza, Marco Turchi, Antonios Anastasopoulos and Matteo Negri	138
<i>A Computational Approach to Poetic Structure, Rhythm and Rhyme</i>	
Rodolfo Delmonte	144
<i>A reevaluation of dependency structure evaluation</i>	
Rodolfo Delmonte	151
<i>Analisi Linguistica e stilostatistica - Uno studio predittivo sul campo</i>	
Rodolfo Delmonte	158
<i>An adaptable morphological parser for agglutinative languages</i>	
Marina Ermolaeva	164
<i>Distributed smoothed tree kernel</i>	
Lorenzo Ferrone and Fabio Massimo Zanzotto	169
<i>Polysemy alternations extraction using the PAROLE SIMPLE CLIPS Italian lexicon</i>	
Francesca Frontini, Valeria Quochi and Monica Monachini	175

<i>Rappresentazione dei concetti azionali attraverso prototipi e accordo nella categorizzazione dei verbi generali. Una validazione statistica.</i>	
Gloria Gagliardi	180
<i>Correcting OCR errors for German in Fraktur font</i>	
Michel Génèreux, Egon W. Stemle, Verena Lyding and Lionel Nicolas	186
<i>Some issues on Italian to LIS automatic translation. The case of train announcements.</i>	
Carlo Geraci, Alessandro Mazzei and Marco Angster	191
<i>ConParoleTue: crowdsourcing al servizio di un Dizionario delle Collocazioni Italiane per Apprendenti (Dici-A)</i>	
Andrea Gobbi and Stefania Spina	197
<i>Making Latent SVMstruct practical for coreference resolution</i>	
Iryna Haponchyk and Alessandro Moschitti	203
<i>Nominal coercion in space: Mass/count nouns and distributional semantics</i>	
Manuela Hürlimann, Raffaella Bernardi and Denis Paperno	208
<i>Part-of-Speech tagging strategy for MIDIA: a diachronic corpus of the Italian language</i>	
Claudio Iacobini, Aurelio De Rosa and Giovanna Schirato	213
<i>Distributional analysis of copredication: Towards distinguishing systematic polysemy from coercion</i>	
Elisabetta Jezek and Laure Vieu	219
<i>Publishing PAROLE SIMPLE CLIPS as linguistic linked open data</i>	
Fahad Khan and Francesca Frontini	224
<i>A preliminary comparison of state-of-the-art dependency parsers on the Italian Stanford Dependency Treebank</i>	
Alberto Lavelli	229
<i>SYMPATHy: Towards a comprehensive approach to the extraction of Italian Word Combinations</i>	
Alessandro Lenci, Gianluca E. Lebani, Sara Castagnoli, Francesca Masini and Malvina Nissim	234
<i>Più l'ascolto e più Mi piace! Social media e radio: uno studio preliminare del successo dei post</i>	
Eleonora Lisi, Emanuele Donati and Fabio Massimo Zanzotto	239
<i>Estimating lexical resources Impact in text-to-text inference tasks</i>	
Simone Magnolini and Bernardo Magnini	244
<i>Parting ways with the partitive view: a corpus-based account of the Italian particle “ne”</i>	
Alice Mariotti and Malvina Nissim	249
<i>On the lexical coverage of some resources on Italian cooking recipes</i>	
Alessandro Mazzei	254
<i>Event Factuality in Italian: Annotation of News Stories from the Ita-TimeBank</i>	
Anne-Lyse Minard, Alessandro Marchetti and Manuela Speranza	260
<i>An English-Italian MWE dictionary</i>	
Johanna Monti	265
<i>ALCIDE: An online platform for the analysis of language and content in a digital environment</i>	
Giovanni Moretti, Sara Tonelli, Stefano Menini and Rachele Sprugnoli	270

<i>Inner speech, dialogue text and collaborative learning in virtual learning communities</i>	
Stefanos Nikiforos and Katia Lida Kermanidis	275
<i>Gli errori di un sistema di riconoscimento automatico del parlato. Analisi linguistica e primi risultati di una ricerca interdisciplinare</i>	
Maria Palmerini and Renata Savy	281
<i>“Il Piave mormorava...”: Recognizing Locations and other Named Entities in Italian Texts on the Great War</i>	
Lucia Passaro and Alessandro Lenci	286
<i>The importance of being sum. Network analysis of a Latin dependency treebank</i>	
Marco Passarotti	291
<i>I-ChatBIT: An intelligent chatbot for the Italian Language</i>	
Arianna Pipitone, Vincenzo Cannella and Roberto Pirrone	296
<i>Two-dimensional wordlikeness effects in lexical organisation</i>	
Vito Pirrelli, Claudia Marzi and Marcello Ferro	301
<i>Toward disambiguating typed predicate-argument structures for Italian</i>	
Octavian Popescu, Ngoc Phuoc An Vo, Anna Feltracco, Elisabetta Jezek and Bernardo Magnini	306
<i>Il corpus Speaky</i>	
Fabio Poroli, Massimiliano Todisco, Michele Cornacchia, Cristina Delogu, Andrea Paoloni and Mauro Falcone	311
<i>Converting the parallel treebank ParTUT in Universal Stanford Dependencies</i>	
Manuela Sanguinetti and Cristina Bosco	316
<i>Developing corpora and tools for sentiment analysis: the experience of the University of Turin group</i>	
Manuela Sanguinetti, Emilio Sulis, Viviana Patti, Giancarlo Ruffo, Leonardo Allisio, Valeria Mus- sa and Cristina Bosco	322
<i>Unsupervised antonym-synonym discrimination in vector space</i>	
Enrico Santus, Qin Lu, Alessandro Lenci and Chu-Ren Huang	328
<i>Methods of textual archive preservation</i>	
Eva Sassolini, Sebastiana Cucurullo and Manuela Sassi	334
<i>Combining unsupervised syntactic and semantic models of thematic fit</i>	
Asad Sayeed and Vera Demberg	339
<i>Deep neural network adaptation for children’s and adults’ speech recognition</i>	
Romain Serizel and Diego Giuliani	344
<i>An Italian corpus for aspect based sentiment analysis of movie reviews</i>	
Antonio Sorgente, Giuseppe Vettigli and Francesco Mele	349
<i>Il Perugia Corpus: una risorsa di riferimento per l’italiano. Composizione, annotazione e valutazione</i>	
Stefania Spina	354
<i>Are quantum classifiers promising?</i>	
Fabio Tamburini	360
<i>Geometric and statistical analysis of emotions and topics in corpora</i>	
Francesco Tarasconi and Vittorio Di Tomaso	365

<i>Corpus ICoN: una raccolta di elaborati di italiano L2 prodotti in ambito universitario</i>	
Mirko Tivosanis	370
<i>Coreference resolution for Italian: Assessing the impact of linguistic components</i>	
Olga Uryupina and Alessandro Moschitti	374
<i>A context based model for Sentiment Analysis in Twitter for the Italian language</i>	
Andrea Vanzo, Giuseppe Castellucci, Danilo Croce and Roberto Basili	379
<i>Semantic role annotation of instrument subjects</i>	
Rossella Varvara and Elisabetta Jezek	384
<i>The Italian module for NooJ</i>	
Simonetta Vietri	389