

Proceedings  
of the  
Fifth Italian Conference  
on  
Computational Linguistics  
CLiC-it 2018

10-12 December 2018, Torino

Editors:

Elena Cabrio  
Alessandro Mazzei  
Fabio Tamburini



aA



© 2018 by AILC - Associazione Italiana di Linguistica Computazionale  
sede legale: c/o Bernardo Magnini, Via delle Cave 61, 38122 Trento  
codice fiscale 96101430229  
email: [info@ai-lc.it](mailto:info@ai-lc.it)

Pubblicazione resa disponibile  
nei termini della licenza Creative Commons  
Attribuzione – Non commerciale – Non opere derivate 4.0



Accademia University Press  
via Carlo Alberto 55  
I-10123 Torino  
[info@aAccademia.it](mailto:info@aAccademia.it)

isbn 978-88-31978-41-5  
[www.aAccademia.it/CLIC\\_2018](http://www.aAccademia.it/CLIC_2018)

Accademia University Press è un marchio registrato di proprietà  
di LEXIS Compagnia Editoriale in Torino srl

## Table of Contents

Preface.....	1
<b>Keynote Talks</b>	
Computational Semantics in Neural Times Johan Bos, University of Groningen, Netherlands .....	3
Disentangling the Thoughts: Latest News in Computational Argumentation Iryna Gurevych, Technische Universität Darmstadt, Germany .....	4
<b>Contributed Papers</b>	
A distributional study of negated adjectives and antonyms Laura Aina, Raffaella Bernardi, Raquel Fernández .....	7
PRET: Prerequisite-Enriched Terminology. A Case Study on Educational Texts Chiara Alzetta, Frosina Koceva, Samuele Passalacqua, Ilaria Torre, Giovanni Adorni.....	14
Distributional Analysis of Verbal Neologisms: Task Definition and Dataset Construction Matteo Amore, Stephen McGregor, Elisabetta Jezek .....	21
Parsing Italian texts together is better than parsing them alone! Oronzo Antonelli, Fabio Tamburini .....	27
“ <i>Buon appetito!</i> ” - Analyzing Happiness in Italian Tweets Pierpaolo Basile, Nicole Novielli.....	34
Long-term Social Media Data Collection at the University of Turin Valerio Basile, Mirko Lai, Manuela Sanguinetti .....	40
Neural Surface Realization for Italian Valerio Basile, Alessandro Mazzei.....	46
Hurtlex: A Multilingual Lexicon of Words to Hurt Elisa Bassignana, Valerio Basile, Viviana Patti .....	51
CoreNLP-it: A UD pipeline for Italian based on Stanford CoreNLP Alessandro Bondielli, Lucia C. Passaro, Alessandro Lenci.....	57
DARC-IT: a DATaset for Reading Comprehension in ITalian Dominique Brunato, Martina Valeriani, Felice Dell’Orletta .....	62
Modelling Italian construction flexibility with distributional semantics: are constructions enough? Lucia Busso, Ludovica Pannitto, Alessandro Lenci .....	68
The SEEMPAD Dataset for Emphatic and Persuasive Argumentation Elena Cabrio, Serena Villata.....	75
Italian Event Detection Goes Deep Learning Tommaso Caselli.....	81
Enhancing the Latin Morphological Analyser LEMLAT with a Medieval Latin Glossary Flavio Massimiliano Cecchini, Marco Passarotti, Paolo Ruffolo, Marinella Testori, Lia Draetta, Martina Fieromonte, Annarita Liano, Costanza Marini, Giovanni Piantanida.....	87

Is Big Five better than MBTI? A personality computing challenge using Twitter data Fabio Celli, Bruno Lepri .....	93
Automatically Predicting User Ratings for Conversational Systems Alessandra Cervone, Enrico Gambi, Giuliano Tortoreto, Evgeny Stepanov, Giuseppe Riccardi.....	99
An efficient Trie for binding (and movement) Cristiano Chesi.....	105
Generalizing Representations of Lexical Semantic Relations Anupama Chingacham, Denis Paperno.....	111
A NLP-based Analysis of Reflective Writings by Italian Teachers Giulia Chiriatti, Valentina Della Gala, Felice Dell'Orletta, Simonetta Montemagni, Maria Chiara Pettenati, Maria Teresa Sagri, Giulia Venturi.....	118
Sentences and Documents in Native Language Identification Andrea Cimino, Felice Dell'Orletta, Dominique Brunato, Giulia Venturi .....	125
Gender and Genre Linguistic profiling: a case study on female and male journalistic and diary prose Eleonora Cocciu, Dominique Brunato, Giulia Venturi, Felice Dell'Orletta .....	131
Conceptual Abstractness: from Nouns to Verbs Davide Colla, Enrico Mensa, Aureliano Porporato, Daniele P. Radicioni .....	137
Effective Communication without Verbs? Sure! Identification of Nominal Utterances in Italian Social Media Texts Gloria Comandini, Manuela Speranza, Bernardo Magnini.....	143
On the Readability of Deep Learning Models: the role of Kernel-based Deep Architectures Danilo Croce, Daniele Rossini, Roberto Basili .....	149
The CHROME Manifesto: integrating multimodal data into Cultural Heritage Resources Francesco Cutugno, Felice Dell'Orletta, Isabella Poggi, Renata Savy, Antonio Sorgente.....	155
Italian in the Trenches: Linguistic Annotation and Analysis of Texts of the Great War Irene De Felice, Felice Dell'Orletta, Giulia Venturi, Alessandro Lenci, Simonetta Montemagni .....	160
Lexicon and Syntax: Complexity across Genres and Language Varieties Pietro Dell'Oglio, Dominique Brunato, Felice Dell'Orletta.....	165
Grammatical class effects in production of Italian inflected verbs Maria De Martino, Azzurra Mancuso, Alessandro Laudanna .....	171
Integrating Terminology Extraction and Word Embedding for Unsupervised Aspect Based Sentiment Analysis Luca Dini, Paolo Curtoni, Elena Melnikova.....	176
A Linguistic Failure Analysis of Classification of Medical Publications: A Study on Stemming vs Lemmatization Giorgio Maria Di Nunzio, Federica Vezzani.....	182
Lexical Opposition in Discourse Contrast Anna Feltracco, Bernardo Magnini, Elisabetta Jezek.....	187
A new Pitch Tracking Smoother based on Deep Neural Networks Michele Ferro, Fabio Tamburini.....	193

Using and evaluating TRACER for an <i>Index fontium computatus of the Summa contra Gentiles</i> of Thomas Aquinas.....	199
Greta Franzini, Marco Passarotti, Maria Moritz, Marco Büchler	
Inter-Annotator Agreement in linguistica: una rassegna critica (ENGLISH Inter-Annotator Agreement in linguistics: a critical review)	
Gloria Gagliardi.....	206
Auxiliary selection in Italian intransitive verbs: a computational investigation based on annotated corpora	
Ilaria Ghezzi, Cristina Bosco, Alessandro Mazzei.....	213
Constructing an Annotated Resource for Part-Of-Speech Tagging of Mishnaic Hebrew	
Emiliano Giovannetti, Davide Albanesi, Andrea Bellandi, Simone Marchi, Alessandra Pecchioli .....	219
Concept Tagging for Natural Language Understanding: Two Decadelong Algorithm Development	
Jacopo Gobbi, Evgeny Stepanov, Giuseppe Riccardi.....	224
The language-invariant aspect of compounding: Predicting compound meanings across languages	
Fritz Günther, Marco Marelli.....	230
From General to Specific : Leveraging Named Entity Recognition for Slot Filling in Conversational Language Understanding	
Samuel Louvan, Bernardo Magnini .....	235
What's in a Food Name: Knowledge Induction from Gazetteers of Food Main Ingredient	
Bernardo Magnini, Vevake Balaraman, Simone Magnolini, Marco Guerini .....	241
La sentiment analysis come strumento di studio del parlato emozionale? (ENGLISH Can sentiment analysis support the study of emotional speech?)	
Paolo Mairano, Enrico Zovato, Vito Quinci.....	247
The iDAI publication: extracting and linking information in the publications of the German Archaeological Institute (DAI)	
Francesco Mambrini.....	253
Source-driven Representations for Hate Speech Detection	
Flavio Merenda, Claudia Zaghi, Tommaso Caselli, Malvina Nissim.....	258
Progettare Chatbot: considerazioni e linee guida (ENGLISH Designing chatobot: analysis and guidelines)	
Eleonora Mollo, Amon Rapp, Dario Mana, Rossana Simeoni.....	264
Advances in Multiword Expression Identification for the Italian language: The PARSEME shared task edition 1.1	
Johanna Monti, Silvio Ricardo Cordeiro, Carlos Ramisch, Federico Sangati, Agata Savary, Veronika Vincze.....	271
PARSEME-IT - Issues in verbal Multiword Expressions identification and classification	
Johanna Monti, Valeria Caruso, Maria Pia Di Buono .....	278
Local associations and semantic ties in overt and masked semantic priming	
Andrea Nadalini, Marco Marelli, Roberto Bottini, Davide Crepaldi.....	283
Online Neural Automatic Post-editing for Neural Machine Translation	
Matteo Negri, Marco Turchi, Nicola Bertoldi, Marcello Federico.....	288
EnetCollect in Italy	
Lionel Nicolas, Verena Lyding, Luisa Bentivogli, Federico Sangati, Johanna Monti, Irene Russo, Roberto Gretter, Daniele Falavigna.....	294

Towards SMT-Assisted Error Annotation of Learner Corpora Nadezda Okinina, Lionel Nicolas .....	299
Towards Personalised Simplification based on L2 Learners' Native Language Alessio Palmero Aprosio, Stefano Menini, Sara Tonelli, Luca Ducceschi, Leonardo Herzog.....	305
Tint 2.0: an All-inclusive Suite for NLP in Italian Alessio Palmero Aprosio, Giovanni Moretti.....	311
MEDEA: Merging Event knowledge and Distributional vEctor Addition Ludovica Pannitto, Alessandro Lenci.....	318
LatInfLexi: an Inflected Lexicon of Latin Verbs Matteo Pellegrini, Marco Passarotti .....	324
Word Embeddings in Sentiment Analysis Ruggero Petrolito, Felice Dell'Orletta.....	330
Identifying Citation Contexts: a Review of Strategies and Goals Agata Rotondi, Angelo Di Iorio, Freddy Limpens .....	335
DialettiBot: a Telegram Bot for Crowdsourcing Recordings of Italian Dialects Federico Sangati, Ekaterina Abramova, Johanna Monti.....	342
Bootstrapping Enhanced Universal Dependencies for Italian Maria Simi, Simonetta Montemagni .....	348
Analysing the Evolution of Students' Writing Skills and the Impact of Neo-standard Italian with the help of Computational Linguistics Rachele Sprugnoli, Sara Tonelli, Alessio Palmero Aprosio, Giovanni Moretti .....	354
<i>Arretium</i> or <i>Arezzo</i> ? A Neural Approach to the Identification of Place Names in Historical Texts Rachele Sprugnoli.....	360
Multi-source Transformer for Automatic Post-Editing Amirhossein Tebbifakhr, Ruchit Agrawal, Matteo Negri, Marco Turchi.....	366
Classifying Italian newspaper text: news or editorial? Pietro Totis, Manfred Stede.....	372
Multi-Word Expressions in spoken language: PoliSdict Daniela Trotta, Michele Stingo, Teresa Albanese, Raffaele Guarasci, Annibale Elia.....	377



Associazione Italiana di  
Linguistica Computazionale



## Preface

On behalf of the Program Committee, a very warm welcome to the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018). This edition of the conference is held in Torino. The conference is locally organised by the University of Torino and hosted into its prestigious main lecture hall “*Cavallerizza Reale*”. The CLiC-it conference series is an initiative of the Italian Association for Computational Linguistics (AILC) which, after five years of activity, has clearly established itself as the premier national forum for research and development in the fields of Computational Linguistics and Natural Language Processing, where leading researchers and practitioners from academia and industry meet to share their research results, experiences, and challenges.

This year CLiC-it received 70 submissions against 64 submissions in 2015, 69 in 2016 and 72 in 2017. The Programme Committee worked very hard to ensure that every paper received at least two careful and fair reviews. This process finally led to the acceptance of 18 papers for oral presentation and 45 papers for poster presentation, with a global acceptance rate of 90% motivated by the inclusive spirit of the conference. The conference is also receiving considerable attention from the international community, with 16 (23%) submissions showing at least one author affiliated to a foreign institution. Regardless of the format of presentation, all accepted papers are allocated 5 pages plus 2 pages for references in the proceedings, available as open access publication. In line with previous editions, the conference is organised around thematic areas managed by one or two area chairs per area.

In addition to the technical programme, this year we are honoured to have as invited speakers internationally recognised researchers as Johan Bos (University of Groningen) and Iryna Gurevych (Technische Universität Darmstadt). We are very grateful to Johan and Iryna for agreeing to share with the Italian community their knowledge and expertise on key topics in Computational Linguistics.

Traditionally, around one half of the participants at CLiC-it are young postdocs, PhD students, and even undergraduate students. As in the previous edition of the conference, we organised a special track called “Research Communications”, encouraging authors of articles published in 2018 at outstanding international conferences in our field to submit short abstracts of their work. Research communications are not published in the proceedings, but are orally presented within a dedicated session at the conference, in order to enforce dissemination of excellence in research.

Moreover, during the conference we award the prize for the best Master Thesis (*Laurea Magistrale*) in Computational Linguistics, submitted at an Italian University between August 1st 2017 and July 31st 2018. This special prize is also endorsed by AILC. We have received 6 candidate theses, which have been evaluated by a special jury. The prize will be awarded at the conference, by a member of the jury.

As last year, we propose a tutorial at the beginning of the conference (Paolo Rosso – Profiling Information in Social Media). We highlight the importance that this kind of opportunities have for young researchers in particular, and we are proud of having made the tutorial attendance free for all registered students.

Even if CLiC-it is a medium size conference, organizing this annual meeting requires major effort from many people. This conference would not have been possible without the dedication, devotion and hard work of the members of the Local Organising Committee, who volunteered their time and energies to contribute to the success of the event. We are also extremely grateful to our Programme Committee members for producing a lot of detailed and insightful reviews, as well as to the Area Chairs who assisted the Programme Chairs in their duties. All these people are named in the following pages. We also want to acknowledge the support from endorsing organisations and institutions and from all of our sponsors, who generously provided funds and services that are crucial for the realisation of this event. Special thanks are also due to the University of Torino for its support in the organisation of the event and for hosting the conference at the main lecture hall “*Cavallerizza Reale*”.

Please join us at CLiC-it 2018 to interact with experts from academia and industry on topics related to Computational Linguistics and Natural Language Processing and to experience and share new research findings, best practices, state-of-the-art systems and applications. We hope that this year’s conference will be intellectually stimulating, and that you will take home many new ideas and methods that will help extend your own research.

Elena Cabrio, Alessandro Mazzei and Fabio Tamburini  
CLiC-it 2018 General Chairs