

ISSN 2499-4553

IJCoL

Italian Journal
of Computational Linguistics

Rivista Italiana
di Linguistica Computazionale

Volume 6, Number 2
december 2020

aA
ccademia
university
press



editors in chief

Roberto Basili

Università degli Studi di Roma Tor Vergata

Simonetta Montemagni

Istituto di Linguistica Computazionale “Antonio Zampolli” - CNR

advisory board

Giuseppe Attardi

Università degli Studi di Pisa (Italy)

Nicoletta Calzolari

Istituto di Linguistica Computazionale “Antonio Zampolli” - CNR (Italy)

Nick Campbell

Trinity College Dublin (Ireland)

Piero Cosi

Istituto di Scienze e Tecnologie della Cognizione - CNR (Italy)

Giacomo Ferrari

Università degli Studi del Piemonte Orientale (Italy)

Eduard Hovy

Carnegie Mellon University (USA)

Paola Merlo

Université de Genève (Switzerland)

John Nerbonne

University of Groningen (The Netherlands)

Joakim Nivre

Uppsala University (Sweden)

Maria Teresa Paziienza

Università degli Studi di Roma Tor Vergata (Italy)

Hinrich Schütze

University of Munich (Germany)

Marc Steedman

University of Edinburgh (United Kingdom)

Oliviero Stock

Fondazione Bruno Kessler, Trento (Italy)

Jun-ichi Tsujii

Artificial Intelligence Research Center, Tokyo (Japan)

Cristina Bosco

Università degli Studi di Torino (Italy)

Franco Cutugno

Università degli Studi di Napoli (Italy)

Felice Dell'Orletta

Istituto di Linguistica Computazionale "Antonio Zampolli" - CNR (Italy)

Rodolfo Delmonte

Università degli Studi di Venezia (Italy)

Marcello Federico

Fondazione Bruno Kessler, Trento (Italy)

Alessandro Lenci

Università degli Studi di Pisa (Italy)

Bernardo Magnini

Fondazione Bruno Kessler, Trento (Italy)

Johanna Monti

Università degli Studi di Sassari (Italy)

Alessandro Moschitti

Università degli Studi di Trento (Italy)

Roberto Navigli

Università degli Studi di Roma "La Sapienza" (Italy)

Malvina Nissim

University of Groningen (The Netherlands)

Roberto Pieraccini

Jibo, Inc., Redwood City, CA, and Boston, MA (USA)

Vito Pirrelli

Istituto di Linguistica Computazionale "Antonio Zampolli" - CNR (Italy)

Giorgio Satta

Università degli Studi di Padova (Italy)

Gianni Semeraro

Università degli Studi di Bari (Italy)

Carlo Strapparava

Fondazione Bruno Kessler, Trento (Italy)

Fabio Tamburini

Università degli Studi di Bologna (Italy)

Paola Velardi

Università degli Studi di Roma "La Sapienza" (Italy)

Guido Vetere

Centro Studi Avanzati IBM Italia (Italy)

Fabio Massimo Zanzotto

Università degli Studi di Roma Tor Vergata (Italy)

Danilo Croce

Università degli Studi di Roma Tor Vergata

Sara Goggi

Istituto di Linguistica Computazionale "Antonio Zampolli" - CNR

Manuela Speranza

Fondazione Bruno Kessler, Trento

Registrazione presso il Tribunale di Trento n. 14/16 del 6 luglio 2016

Rivista Semestrale dell'Associazione Italiana di Linguistica Computazionale (AILC)
© 2020 Associazione Italiana di Linguistica Computazionale (AILC)



Associazione Italiana di
Linguistica Computazionale



direttore responsabile
Michele Arnese

isbn 9791280136626

Accademia University Press
via Carlo Alberto 55
I-10123 Torino
info@aAccademia.it
www.aAccademia.it/IJCoL_6_2



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

CONTENTS

Editorial Note <i>Roberto Basili, Simonetta Montemagni</i>	7
Embeddings-based detection of word use variation in Italian newspapers <i>Michele Cafagna, Lorenzo De Mattei, Malvina Nissim</i>	9
Analyzing Gaussian distribution of semantic shifts in Lexical Semantic Change Models <i>Pierluigi Cassotti, Pierpaolo Basile, Marco de Gemmis, Giovanni Semeraro</i>	23
Linguistically-driven Selection of Difficult-to-Parse Dependency Structures <i>Chiara Alzetta, Felice Dell'Orletta, Simonetta Montemagni, Giulia Venturi</i>	37
ETC-NLG: End-to-end Topic-Conditioned Natural Language Generation <i>Ginevra Carbone, Gabriele Sarti</i>	61
Lessons Learned from EVALITA 2020 and Thirteen Years of Evaluation of Italian Language Technology <i>Lucia C. Passaro, Maria Di Maro, Valerio Basile, Danilo Croce</i>	79
EVALITA4ELG: Italian Benchmark Linguistic Resources, NLP Services and Tools for the ELG Platform <i>Viviana Patti, Valerio Basile, Cristina Bosco, Rossella Varvara, Michael Fell, Andrea Bolioli, Alessio Bosca</i>	105

Editorial Note

Roberto Basili*
Università di Roma, Tor Vergata

Simonetta Montemagni**
ILC - CNR

We are pleased to introduce the second issue of the sixth year of the *Italian Journal of Computational Linguistics* (IJCoL). During the past semester, IJCoL has been awarded "Class A" by ANVUR (Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca - National Agency for the Evaluation of the University System and Academic Research), placing it at the top of the ranking of Scientific Journals for Area 10 "Antiquities, philology, literary studies, art history", Disciplinary Sector 10/G1 "Glottology and Linguistics".

Thanks to this rating, introduced for the scientific fields identified as "not bibliometric" (that is, where citation analysis is not appropriate), IJCoL is now recognised as an excellent journal at national and international levels. This is an important achievement for researchers working in the humanistic area of computational linguistics in Italy. We are also working towards having IJCoL indexed in international citation databases such as Scopus and Web of Science: this goal is particularly important for the bibliometric side of the Italian computational linguistics.

We take here the opportunity to thank the members of the community who have contributed to the journal by reviewing papers, those who have submitted to the journal relevant research papers since the beginning as well as current members of the Editorial and Advisory Boards. A very special thanks goes to the members of the Editorial Office, for their continuous work and commitment in providing an efficient editorial process able to maintaining high standards within our publication.

This issue is a miscellaneous volume which reports original findings achieved in the currently investigated lines of research within the national computational linguistics community. The contributions collected in the volume cover recent and fruitful developments in computational linguistics research, ranging from the computational modelling of lexical semantic change from different perspectives (both diachronic and synchronic), to unsupervised methods for identifying difficult-to-parse syntactic constructions and Natural Language Generation, to the EVALITA evaluation campaign specifically devoted to Natural Language Processing and Speech tools for Italian, its past and current achievements.

The contribution by Cafagna and colleagues opens the volume. The paper investigates how words are used differently in two Italian newspapers at opposite ends of the political spectrum by training embeddings on one newspaper's corpus, updating the weights on the second one, and observing vector shifts. The results of different types of analysis (top-down vs bottom-up) are reported, with interesting results. Despite the

* Dept. of Enterprise Engineering - Via del Politecnico 1, 00133 Roma
E-mail: basili@info.uniroma2.it

** Istituto di Linguistica Computazionale "A. Zampolli", CNR - Via Moruzzi 1, 56124 Pisa
E-mail: simonetta.montemagni@ilc.cnr.it

analysis is specific to the data taken into account, the proposed method can also be seen as a blueprint for similar studies.

The second paper, by Cassotti *et al.*, tackles the topic of lexical semantic change detection from a diachronic perspective. In particular, it reports the evaluation results of graded Lexical Semantic Change Models using thresholds based on the Gaussian distribution of the cosine similarity. The models considered range from Dynamic Word Embeddings, Temporal Random Indexing, Temporal Referencing, to OP-SGNS and Temporal Word Embeddings with a Compass. The evaluation was performed using datasets coming from SemEval-2020 Task 1 Subtask 1 and DIACR-Ita. Results obtained with Gaussian thresholds achieve state-of-the-art performance in English, German, Swedish and Italian.

The contribution by Alzetta and colleagues illustrates a novel methodology meeting the goals of both quantifying the reliability of automatically generated dependency relations without using gold data, and investigating which are the linguistic constructions negatively affecting the parser performance. The results of two experiments, aimed at assessing the degree of parsing difficulty across different dependency relation types, and different instances of the same relation, are reported. They demonstrate that the proposed methodology is able to identify difficult-to-parse dependency relations without relying on gold data and by taking into account a variety of intertwined linguistic factors. These findings pave the way to novel applications, both for defining new dependency parsing evaluation metrics and towards the creation of challenge sets.

Carbone and Sarti present ETC-NLG, an end-to-end method leveraging topic modeling annotations on unlabeled text corpora to generate topic-conditioned sentences in natural language. The method is aimed at dealing with insufficient labeled training data and can be used to produce high quality conditioned text when provided with suitable topic models and parameters that balance generation fluency and conditioning strength. The method's effectiveness was tested in a low-resource setting for Italian and a comparative evaluation of ETC-NLG for Italian and English using a parallel corpora was performed.

The last two papers in the volume focus on the *Evaluation Campaign of Natural Language Processing and Speech Tools for Italian*, EVALITA. Passaro *et al.* provide a summary of the 7th EVALITA campaign (2020) which included 14 different tasks belonging to five research areas, namely: (i) Affect, Hate, and Stance, (ii) Creativity and Style, (iii) New Challenges in Long-standing Tasks, (iv) Semantics and Multimodality, (v) Time and Diachrony. The paper provides a description of the tasks and the key findings from the analysis of participant outcomes, as well as a detailed analysis of the evaluation of tasks across the past seven editions over the last 13 years. The paper by Patti and colleagues closes the volume, describing the EVALITA4ELG project, whose aim is collecting the resources released as benchmarks for the EVALITA campaigns, and making them easily accessible through the European Language Grid platform. The collection of resources is integrated with systems and baselines as a pool of web services with a common interface, deployed on a dedicated hardware infrastructure.

After this synthetic view of the papers in this issue, we leave the reader the pleasure to navigate across the valuable pages of the volume.