

ISSN 2499-4553

# IJCoL

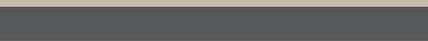
Italian Journal  
of Computational Linguistics

Rivista Italiana  
di Linguistica Computazionale

Volume 1, Number 1  
december 2015

Emerging Topics at the First Italian Conference  
on Computational Linguistics

**aA**ccademia  
university  
press



editors in chief

**Roberto Basili**

Università degli Studi di Roma Tor Vergata (Italy)

**Simonetta Montemagni**

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

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)

editorial board

**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)

editorial office

**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 in corso presso il Tribunale di Trento

Rivista Semestrale dell'Associazione Italiana di Linguistica Computazionale (AILC)  
© 2015 Associazione Italiana di Linguistica Computazionale (AILC)



direttore responsabile  
Michele Arnese

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



ISSN 2499-4553  
ISBN 978-88-99200-63-3

[www.aAccademia.it/IJCoL\\_01](http://www.aAccademia.it/IJCoL_01)

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



## Emerging Topics at the First Italian Conference on Computational Linguistics

*a cura di*  
Roberto Basili, Alessandro Lenci,  
Bernardo Magnini, Simonetta Montemagni

### CONTENTS

Nota Editoriale <i>Roberto Basili, Alessandro Lenci, Bernardo Magnini, Simonetta Montemagni</i>	7
Distributed Smoothed Tree Kernel <i>Lorenzo Ferrone, Fabio Massimo Zanzotto</i>	17
An exploration of semantic features in an unsupervised thematic fit evaluation framework <i>Asad Sayeed, Vera Demberg, and Pavel Shkadzko</i>	31
When Similarity Becomes Opposition: Synonyms and Antonyms Discrimination in DSMs <i>Enrico Santus, Qin Lu, Alessandro Lenci, Chu-Ren Huang</i>	47
Temporal Random Indexing: A System for Analysing Word Meaning over Time <i>Pierpaolo Basile, Annalina Caputo, Giovanni Semeraro</i>	61
Context-aware Models for Twitter Sentiment Analysis <i>Giuseppe Castellucci, Andrea Vanzo, Danilo Croce, Roberto Basili</i>	75
Geometric and statistical analysis of emotions and topics in corpora <i>Francesco Tarasconi, Vittorio Di Tomaso</i>	91
Il ruolo delle tecnologie del linguaggio nel monitoraggio dell'evoluzione delle abilità di scrittura: primi risultati <i>Alessia Barbagli, Pietro Lucisano, Felice Dell'Orletta, Simonetta Montemagni, Giulia Venturi</i>	105
CLaSSES: a new digital resource for Latin epigraphy <i>Irene De Felice, Margherita Donati, Giovanna Marotta</i>	125



## Nota Editoriale

Roberto Basili\*  
Università di Roma, Tor Vergata

Alessandro Lenci\*\*  
Università di Pisa

Bernardo Magnini†  
Fondazione Bruno Kessler, Trento

Simonetta Montemagni‡  
ILC-CNR, Pisa

Siamo felici di introdurre il nuovo *Italian Journal of Computational Linguistics* (IJCoL), la *Rivista Italiana di Linguistica Computazionale*. La rivista nasce e viene pubblicata dalla neo-costituita “Associazione Italiana di Linguistica Computazionale” (AILC - [www.ai-lc.it](http://www.ai-lc.it)) e, assieme alla conferenza annuale CLIC-it (“Italian Conference on Computational Linguistics”) e a EVALITA, la campagna di valutazione per le tecnologie del linguaggio per la lingua italiana scritta e parlata, costituisce uno degli strumenti principali al servizio della comunità italiana per la promozione e per la diffusione della ricerca nel campo della linguistica computazionale affrontata da prospettive diverse e complementari.

L’AILC nasce in un contesto italiano in cui esistono da tempo diverse realtà associative che operano nell’ambito delle scienze del linguaggio. Alcune di esse hanno nella linguistica il loro ambito primario, come la “Società Italiana di Glottologia” (SIG), la “Società di Linguistica Italiana” (SLI), l’ “Associazione Italiana delle Scienze della Voce” (AISV) e l’ “Associazione Italiana di Linguistica Applicata” (AITLA). Altre invece hanno una vocazione più spiccatamente informatica, come l’ “Associazione Italiana di Intelligenza Artificiale” (AI\*IA), o collocano il linguaggio all’interno di più ampie prospettive tematiche, come l’ “Associazione per l’Informatica Umanistica e la Cultura Digitale” (AIUCD) e l’ “Associazione Italiana di Scienze Cognitive” (AISC). Anche le riviste italiane in ambito linguistico non mancano. Tra queste, possiamo citare *Lingue e Linguaggio*, *Studi e Saggi Linguistici* e l’ *Italian Journal of Linguistics*. La rivista *Intelligenza Artificiale* ha inoltre spesso ospitato articoli e numeri tematici sul trattamento automatico del linguaggio.

In questo panorama così ricco e articolato, la domanda spontanea è se fosse necessario creare un’associazione dedicata alla linguistica computazionale. La nostra risposta è, senza alcuna esitazione, un sì forte e convinto. Il motivo fondamentale è che la linguistica computazionale presenta caratteri specifici che la rendono comunque autonoma rispetto alle aree ad essa limitrofe. Diversamente dalle associazioni linguistiche, l’AILC mette al centro dei suoi interessi l’uso dei metodi quantitativi e computazionali

---

\* Dipartimento di Ingegneria dell’Impresa - Via del Politecnico 1, 00133 Rome.  
E-mail: [basili@info.uniroma2.it](mailto:basili@info.uniroma2.it)

\*\* Dipartimento di Filologia, Letteratura e Linguistica - Via Santa Maria 36, 56126 Pisa.  
E-mail: [alessandro.lenci@unipi.it](mailto:alessandro.lenci@unipi.it)

† Fondazione Bruno Kessler - Via Sommarive 18, 38122 Povo, Trento. E-mail: [magnini@fbk.eu](mailto:magnini@fbk.eu)

‡ Istituto di Linguistica Computazionale “Antonio Zampolli” (ILC-CNR) - Via Moruzzi 1, 56124, Pisa.  
E-mail: [simonetta.montemagni@ilc.cnr.it](mailto:simonetta.montemagni@ilc.cnr.it)

per lo studio del linguaggio e lo sviluppo di modelli e tecniche per il trattamento della lingua. Al tempo stesso per AILC è il linguaggio, in tutte le sue manifestazioni, l'oggetto prioritario di ricerca differenziandosi così da quelle realtà che invece collocano il linguaggio nei più ampi domini della modellazione computazionale dell'intelligenza, delle scienze cognitive o dell'informatica applicata alle discipline umanistiche. Autonomia non significa chiusura o separazione. Siamo anzi convinti che AILC dovrà e saprà dialogare con tutte le altre associazioni e realtà interessate al linguaggio e alle lingue naturali. Al tempo stesso, rivendichiamo però un spazio di specificità della linguistica computazionale, che ha bisogno dunque dei suoi spazi di rappresentanza.

Il nuovo *Italian Journal of Computational Linguistics* colma un duplice vuoto, sul versante nazionale e internazionale. Nel panorama editoriale della comunità scientifica italiana, dopo l'esperienza di *Linguistica Computazionale*, fondata nel 1981 da Antonio Zampolli e non più pubblicata dal 2006, è venuto a mancare del tutto un forum autorevole in cui rappresentare le diverse anime della linguistica computazionale in Italia. *Linguistica Computazionale* era espressione di una singola istituzione, l'Istituto di Linguistica Computazionale del CNR, storicamente il primo centro dedicato alla linguistica computazionale a livello nazionale. Oggi, come testimoniato dalla fondazione dell'AILC che riunisce la comunità italiana che opera nel settore, il panorama in Italia è profondamente cambiato, i gruppi di ricerca che si occupano di linguistica computazionale sono numerosi, si estendono su tutto il territorio nazionale e operano sia nell'area umanistica che in quella informatica. Ciò ha reso ancora più urgente la necessità di una rivista che fosse l'espressione della pluralità di voci all'interno della neo-costituita associazione. Questa mancanza è tanto più evidente se consideriamo l'alta reputazione e la visibilità internazionale che la ricerca italiana si è guadagnata nel nostro campo. Sempre sul versante nazionale, IJCoL colma un vuoto evidente ormai da troppo tempo rispetto a iniziative analoghe in altri paesi europei. Pensiamo, ad esempio, alla tradizione e al ruolo che hanno riviste come *Traitement Automatique des Langues* (TAL) per la comunità francese, *Procesamiento del Lenguaje Natural* (PLN) per la comunità spagnola, o *Journal for Language Technology and Computational Linguistics* (JLCL) per quella tedesca. Sul versante internazionale, IJCoL intende contribuire a rafforzare la presenza di riviste del settore della linguistica computazionale, al momento ancora esigua.

Vorremmo che IJCoL fosse riconosciuto come uno strumento per la pubblicazione di risultati di qualità e ottenuti con rigore metodologico, anche quando questi contributi faticano a trovare spazi adeguati in sedi internazionali, vuoi per la scarsità di opportunità in campo editoriale nel nostro settore, vuoi perché non sempre risultati di rilievo ottenuti per la lingua italiana sono valorizzati sufficientemente a livello internazionale. Vorremmo uno spazio di discussione aperto, particolarmente ai contributi di giovani ricercatori, in cui poter riportare esperienze, risultati teorici e sperimentali in uno spirito di confronto continuo, avendo consapevolezza della complessità delle sfide scientifiche e tecnologiche che la linguistica computazionale è chiamata oggi ad affrontare.

Con questo spirito, la rivista intende coprire un ampio spettro di temi che ruotano attorno a linguaggio e computazione affrontato da prospettive diverse che includono ma non si limitano a: trattamento automatico del linguaggio (scritto e parlato), apprendimento automatico del linguaggio, modelli computazionali del linguaggio, della cognizione e della variazione linguistica, acquisizione di conoscenza, costruzione di risorse linguistiche, sviluppo di infrastrutture per l'interoperabilità e l'integrazione di risorse e tecnologie linguistiche, per arrivare a temi con una forte valenza applicativa come ad esempio Information Extraction, Question Answering, sommarizzazione automatica e traduzione automatica. In particolare, la rivista intende proporsi come forum aggiornato di discussione della comunità dei ricercatori che operano nel settore della

linguistica computazionale da prospettive diverse, anche con l'obiettivo di creare un ponte tra i risultati che emergono nelle diverse aree del trattamento automatico del linguaggio e altre discipline, da quelle che con la linguistica computazionale condividono l'oggetto di studio, ovvero le lingue e il linguaggio nelle loro varie manifestazioni (ad esempio, la linguistica, la linguistica italiana, la sociolinguistica, la dialettologia, la filologia), a quelle che con essa condividono metodi di elaborazione e analisi come l'informatica e l'intelligenza artificiale, per arrivare a quelle che possono beneficiare di risorse e tecnologie linguistiche per l'accesso e la gestione delle proprie basi documentali. Particolare attenzione sarà dedicata da un lato alle neuroscienze cognitive, nelle quali la modellazione computazionale ha da sempre un ruolo centrale, e dall'altro al contributo della linguistica computazionale all'interno del più ampio settore delle Digital Humanities, di antica tradizione a livello nazionale ed oggi in pieno sviluppo.

Il bacino d'utenza della rivista è rappresentato dalla comunità scientifica di ricerca della linguistica computazionale in ambito sia accademico che industriale a livello nazionale e internazionale, e potrà anche includere potenziali "stakeholders" interessati ad applicazioni basate su risorse e tecnologie per il trattamento automatico del linguaggio.

La struttura scientifico-editoriale della rivista è articolata come segue:

- la Direzione scientifica, composta da due Co-Direttori rappresentanti delle anime umanistica e informatica della linguistica computazionale italiana, che avrà il compito di verificare la qualità scientifica, il rispetto degli obiettivi e la coerenza della linea editoriale della rivista e si occuperà della sua promozione a livello nazionale e internazionale;
- il Comitato Scientifico, composto da rappresentanti della comunità nazionale e internazionale della linguistica computazionale e selezionati in qualità di esperti delle principali aree di interesse della rivista. La funzione del Comitato Scientifico sarà di indirizzo e supervisione della linea editoriale della rivista;
- il Comitato Editoriale, composto da rappresentanti della comunità nazionale della linguistica computazionale afferente all'AILC e delle diverse aree di competenza, con la funzione di definire la politica editoriale della rivista, supervisionare la valutazione di merito degli articoli proposti e di coordinare l'attività editoriale;
- la Segreteria di Redazione, composta da rappresentanti di diverse istituzioni coinvolte in AILC, che fornirà un supporto operativo al Comitato Editoriale.

IJCoL nasce come rivista peer-reviewed con cadenza semestrale e gratuitamente consultabile e scaricabile on-line nel rispetto dei requisiti dell'*Open Access*, una scelta che vuole favorire il più largo accesso possibile da parte di tutti gli interessati, in quell'ottica di inclusione che guida l'AILC. L'obiettivo a medio-lungo termine è di avere la rivista collocata in fascia "A" per le aree scientifico-disciplinari rilevanti della classificazione ANVUR a livello nazionale (ovvero, L-LIN/01, INF/01, ING-INF/05), e indicizzata nei principali database internazionali per i settori coperti dalla rivista (tra questi, Scopus Bibliographic Database, ERIH Plus, Google Scholar, Web of Science).

Siamo consapevoli che il compito che ci aspetta non è semplice. I modi della ricerca scientifica stanno rapidamente cambiando, e per una rivista nuova non sarà facile guadagnare e mantenere prestigio e autorevolezza. La strada per questi obiettivi ambiziosi passa necessariamente dall'impegno e dalla passione di chi dovrà guidare la realizzazione della rivista, ma anche dal coinvolgimento attivo della comunità scientifica interessata, da varie prospettive, alla linguistica computazionale e al trattamento automatico del linguaggio.

Questo volume è il primo di una serie con cui la rivista seguirà la ricerca e i risultati principali della comunità italiana e internazionale della linguistica computazionale. Nel primo numero, abbiamo deciso di concentrarci sui migliori articoli firmati da giovani ricercatori della Conferenza CLIC-it 2014, tenutasi a Pisa il 9 e 10 dicembre 2014. Questi articoli sono stati selezionati tra tutte le aree tematiche della conferenza, in modo da essere rappresentativi dei vari interessi scientifici della nostra comunità, in particolare dei suoi più giovani protagonisti. Gli articoli di questo numero, selezionati attraverso un processo di peer-review, sono stati valutati ulteriormente durante i lavori della Conferenza: questo processo ha portato all'assegnazione dei premi di "Best Young Paper" e "Distinguished Young Papers". Gli autori insigniti di tali riconoscimenti sono stati invitati a sottomettere una versione rivista ed estesa del loro contributo alla conferenza, che è stato oggetto di un'ulteriore valutazione. Il risultato è un numero della rivista che rappresenta linee di ricerca originali e innovative all'interno della comunità della linguistica computazionale italiana, ma non soltanto.

I lavori qui raccolti possono essere ripartiti in quattro aree tematiche generali. In una prima area collochiamo il lavoro di Ferrone e Zanzotto, il cui obiettivo principale è la modellizzazione matematica di informazioni linguistiche di livello lessicale o frasale. Questo lavoro discute come l'integrazione di rappresentazioni grammaticali distribuite, in genere veicolate tramite i cosiddetti "tree kernel", con modelli compositivi possa essere realizzata in processi di apprendimento automatico di tipo linguistico. Il lavoro propone un paradigma unificato che enfatizza al contempo la conoscenza grammaticale e lessicale così come l'algoritmica induttiva ed una rigorosa modellazione matematica.

In un secondo gruppo, troviamo lavori sulla *semantica lessicale*, nella prospettiva specifica dei modelli di rappresentazione vettoriale, ispirati alla ricerca nella semantica distribuzionale. Il lavoro di Sayeed e dei suoi colleghi esplora l'uso di rappresentazioni tensoriali nello studio del cosiddetto "thematic fit", ovvero il grado di congruenza di un argomento rispetto ai vincoli semantici imposti dall'evento espresso da un predicato. Un elemento originale di questo lavoro è la costruzione di uno spazio vettoriale che integra informazione sui ruoli semantici ottenuta attraverso SENNA, un'architettura di *deep learning* per il *semantic role labeling*.

Il lavoro di Santus et al. studia metodi distribuzionali nella modellazione della opposizione semantica tra i sensi lessicali, fenomeno particolarmente complesso per i modelli distribuzionali. Il lavoro propone APant, una misura di (dis)similarità basata sull'assunzione che gli opposti sono simili dal punto di vista distribuzionale ma esprimono differenze tra loro in almeno una delle dimensioni semantiche salienti. Nell'esauritiva analisi sperimentale discussa nell'articolo, si mostra come APant migliori le misure di metodi precedentemente pubblicati nel task di riconoscimento di antonimi.

Il lavoro di Basile et al. propone l'uso del Random Indexing (RI) nello studio della evoluzione diacronica del senso delle parole in corpora che coprono ampi periodi storici. Nell'articolo viene presentato il metodo di Temporal Random Indexing per la acquisizione di spazi di parole dipendenti dal tempo e di esso viene discussa la sperimentazione su due corpora rappresentativi di periodi diversi: una collezione di libri in italiano e i lavori scientifici in lingua inglese nell'area della linguistica computazionale.

Un terzo gruppo di lavori si focalizza sull'*applicazione dell'elaborazione della lingua nel riconoscimento automatico delle opinioni e delle emozioni* nei testi e, in particolare, nelle *Reti Sociali*.

Il lavoro di Castellucci e dei suoi colleghi discute un approccio basato sull'apprendimento strutturato nel riconoscimento di opinioni nei messaggi su Twitter. Qui vengono integrate tecniche di semantica distribuzionale e di apprendimento basato

su “kernel” all’interno di un metodo di classificazione delle opinioni nei microblog sensibile al contesto attraverso una formulazione markoviana di una Support Vector Machine. La sperimentazione condotta su Italiano ed Inglese mostra come il modello migliori i risultati di approcci non strutturati precedentemente proposti in letteratura.

Metodi quantitativi applicati alla semantica lessicale caratterizzano anche l’applicazione dell’elaborazione linguistica al riconoscimento di tematiche ed emozioni negli scenari delle Social TV, come discusso nel lavoro di Tarasconi e Di Tomaso. Essi propongono l’analisi delle corrispondenze multiple come strumento per lo studio delle dipendenze tra temi di discussione ed emozioni. La valutazione sperimentale discute dati estratti da Twitter tra l’ottobre 2013 ed il febbraio 2014, dimostrando l’efficacia e la relativa semplicità di applicazione del metodo.

L’ultima sezione del volume include interessanti esperienze di applicazione di metodi e tecniche della linguistica computazionale nell’ambito di discipline umanistiche, quali la pedagogia sperimentale e lo studio delle lingue classiche.

Il lavoro di Barbagli et al. è focalizzato sull’uso di tecnologie del linguaggio per l’analisi dei processi di apprendimento. Il contributo riporta i primi risultati di uno studio interdisciplinare a cavallo tra linguistica computazionale, linguistica e pedagogia sperimentale finalizzato al monitoraggio dell’evoluzione del processo di apprendimento della lingua italiana come L1. Tale studio condotto con strumenti di annotazione linguistica automatica ha portato all’identificazione di un insieme di tratti caratterizzanti l’evoluzione del processo di apprendimento linguistico, con potenziali e interessanti ricadute applicative sul versante scolastico ed educativo.

Chiude il volume l’articolo di De Felice et al. che illustra la progettazione e lo sviluppo di un’innovativa risorsa digitale per l’epigrafia latina, contenente un corpus di iscrizioni latine annotato con informazioni di varia natura (linguistiche, sociolinguistiche e metalinguistiche). L’articolo illustra l’annotazione della prima macrosezione del corpus relativa a iscrizioni latine del periodo arcaico, che crea i presupposti per raffinate analisi sociolinguistiche del latino preclassico di natura qualitativa e quantitativa a partire da attestazioni epigrafiche.

La breve vista d’insieme sin qui discussa non può coprire i così tanti aspetti di interesse dei lavori citati, e lascia al lettore l’onere, unito speriamo al piacere, di approfondirli direttamente negli articoli in questo volume. In ogni caso, essi ci mostrano con chiarezza l’ampiezza e la granularità dei contributi stimolati dalla prima “Conferenza italiana di Linguistica Computazionale”, CLIC-it 2014. Come suo risultato diretto, dunque, questo numero della rivista è un ulteriore segno tangibile del potenziale esibito regolarmente dalla comunità italiana, che contribuisce in modo significativo alla dimensione internazionale della ricerca in linguistica computazionale.

## Editorial Note Summary

We are pleased to announce the new *Italian Journal of Computational Linguistics* (IJCoL), in Italian *Rivista Italiana di Linguistica Computazionale*. The journal is published by the newly founded Italian Association of Computational Linguistics (AILC - [www.ai-lc.it](http://www.ai-lc.it)). Together with the annual conference CLIC-it (“Italian Conference on Computational Linguistics”) and the EVALITA evaluation campaign specifically devoted to Natural Language Processing and Speech tools for Italian, this journal is intended to meet the need for a national and international forum for the promotion and dissemination of high-level original research in the field of Computational Linguistics (CL).

The journal intends to fill a twofold gap, at the national and international levels. After the journal *Linguistica Computazionale* founded in 1981 by Antonio Zampolli and no longer published since 2006, Italy needed an authoritative forum for researchers working in CL from different and complementary perspectives. Today, the Italian Association for Computational Linguistics brings together the Italian community of CL researchers: the research groups working in this area are numerous, extend over the entire national territory, operate in both academic and industrial environments, in humanistic and/or computer science departments. In this context, a journal which was the expression of the plurality of voices within the newly founded Italian association was urgently needed. IJCoL aims at playing the role of journals like *Traitement Automatique des Langues* (TAL) for the French community, or *Procesamiento of Lenguaaje Natural* (PLN) for the Spanish community, or *Journal for Computational Linguistics and Language Technology* (JLCL) for the German one. This lack is even more evident if we consider the high reputation and visibility gained by Italian CL research at the international level. On such a front, IJCoL aims at increasing the still low presence of journals in the area of Computational Linguistics.

We would like IJCoL to publish the results of high-quality methodologically-sound research, which sometimes is struggling to find adequate space in international fora, due either to the limited number of editorial possibilities or to the fact that results obtained for the Italian language are not always properly valued at the international level. We would like IJCoL to be an open space for discussion, particularly by young researchers bringing in experiences, theoretical and experimental results in a continuous dialogue, being aware of the complexity of the scientific and technological challenges that CL is called to face today.

IJCoL intends to cover a broad spectrum of topics related to natural language and computation tackled from different perspectives, including but not limited to: natural language and speech processing, computational natural language learning, computational modelling of language and language variation, linguistic knowledge acquisition, corpus development and annotation, design and construction of computational lexicons, up to more applicative perspectives such as information extraction, ontology engineering, summarization, machine translation and, last but not least, digital humanities. In particular, a central aim of the journal will be to provide a channel of communication among researchers from multiple perspectives, by bridging the gap between the results emerging in the different areas of natural language processing and other disciplines, ranging from theoretical or descriptive linguistics, cognitive psychology, philosophy, philology or neuroscience and computer science.

The intended audience of the journal typically includes academic and industrial researchers in the areas listed above, but also “stakeholders” such as educators, public administrators and all potential users interested in applications making use of linguistic technologies.

The *Italian Journal of Computational Linguistics* will be an open-access peer-reviewed journal published online twice a year; each volume is expected to be around 120 pages. The journal will alternate miscellaneous volumes and special issues aimed at showcasing research focused on particularly crucial topics. In addition to full articles, the journal will also publish shorter notes and book reviews.

IJCoL is guided by different boards as detailed below:

- two Editors in Chief, representing the humanistic and computer science sides of Italian CL;
- the Advisory Board, which includes distinguished scholars drawn from leading CL research groups around the world selected as experts of hot areas of CL research;
- the Editorial Board, including representatives of the Italian national CL community and of different competence areas;
- the Editorial Office.

The first volume of the journal opens the series that we will dedicate to monitor the research and main achievements of the Italian and international CL community. As a starting point, we decided to focus on the best papers of the CLIC-it 2014 Conference held in December 2014 in Pisa, along two major motivations. First, the research work involved by this choice was inherently representative of the entire community, with its interests, major paradigms and achievements. Second, the papers, early selected on the basis of the CLIC-it 2014 peer-review, have been further evaluated, at the Conference, as candidates for the best paper award and their revised versions have undergone a second round of reviewing. For the variety of topics covered and for the general quality of the papers, we can say that the volume successfully sheds light on several interesting active research trends and contributes to their main challenges. The works here collected can be grouped into four major areas, sketched below.

*Mathematical modeling of linguistic information.* The paper by Ferrone and Zanzotto focuses on the mathematical modeling of linguistic information at the sentence and lexical levels. In particular, it discusses how the integration of grammatical representations supporting specific kernels, the so-called “tree kernels”, with compositionality operators can be effectively applied in computational natural language learning. The proposed rich mathematical formalization emphasizes the role of grammatical and lexical knowledge within a unifying inductive process.

*Distributional Semantics.* This second group gathers contributions whose major focus is on lexical semantics as studied within the light of vector space models, inspired by research in Distributional Semantics. The work by Sayeed et al. explores tensor based representations in the study of so-called “thematic fit”, i.e. the strength by which an entity fits a thematic role in the semantic frame of an event. The adoption of a strict semantic view in the unsupervised acquisition of a distributional space (here called SDDM) provides a promising complementary alternative to existing methods based on syntactic information. The study is based on SENNA, a *deep learning* based architecture for *semantic role labeling*.

The work by Santus et al. explores distributional methods for the study of the semantic opposition between lexical senses, representing a complex phenomenon for distributional models. The work discusses APAnt, a (dis)similarity measure, assuming that opposites can be distributionally similar but must be different from each other in

at least one salient dimension of meaning. In an extensive evaluation discussed in the paper, APAnt is shown to outperform existing baselines in an antonym retrieval task.

The work by Basile and colleagues focuses on the use of Random Indexing (RI) for studying the temporal evolution of word senses over corpora covering long time periods. Interestingly, RI supports a unified representation of vectors for different word distributions that can be acquired over different time spans. In the paper, the Temporal Random Indexing method for building WordSpaces that accounts for temporal information is correspondingly presented and experimented over two corpora: a collection of Italian books and English scientific papers about CL.

*Automatic recognition of opinions and emotions in corpora and Social Networks.* A third group of papers clusters around applications of language analysis to the automatic recognition of opinions and emotions in corpora and Social Networks. In particular, the paper by Castellucci et al. focuses on a structured learning approach for the recognition of opinions over microblogging messages of Twitter. Methods for distributional vector-based lexical representations and kernel-based learning are integrated within a context-aware opinion classification method. The task of recognizing the polarity of a message is here mapped into a tweet sequence labeling task. A Markovian formulation of the Support Vector Machine discriminative approach is applied and reported empirical validation shows how it outperforms existing methods for polarity detection over Italian and English data.

Quantitative methods for lexical semantics also characterize the application of complex language processing chains to the recognition of topics and emotions in Social TV scenarios, as discussed in the paper by Tarasconi and Di Tomaso. They propose Multiple Correspondence Analysis as a tool for studying how audiences share their feelings and representing these similarities in a sound and compact manner. The reported empirical investigation discusses Twitter data extracted between October 2013 and February 2014 showing the effectiveness and viability of the method.

*Application of language processing methods in Digital Humanities.* The last group of papers focuses on the application of natural language processing methods in digital humanities, such as education, epigraphy and sociolinguistics. The paper by Barbagli et al. shows that nowadays the use of language technologies can be successfully extended to the study of learning processes. The paper reports some first results of an interdisciplinary study, as part of a broader experimental pedagogy project, aimed at monitoring the evolution of the learning process of the Italian language based on a corpus of written productions by students, which has been analyzed with automatic linguistic annotation and knowledge extraction tools. Achieved results are very promising and led to the identification of linguistic features qualifying the evolution of language acquisition.

The paper by De Felice and colleagues presents CLaSSES (Corpus for Latin Sociolinguistic Studies on Epigraphic textS), an annotated corpus aimed at (socio)linguistic research on Latin inscriptions: in particular, it illustrates the first macro-section of CLaSSES, including inscriptions of the archaic and early periods (CLaSSES I). Annotated with linguistic, extra- and meta-linguistic features, the corpus can be used to perform quantitative and qualitative variationist analyses on Latin epigraphic texts: it allows the user to analyze spelling (and possibly phonetic-phonological) variants and to interpret them with reference to time, location and text type.

Our synthetic and overall view does not exhaust the wide range of issues explored by the papers and leaves the reader the burden, and, hopefully, the pleasure, discover them in the rest of the volume. However, it clearly shows the width and depth of the contributions produced by the CLIC-it 2014 Conference. As a by product of its lively and vital activity, this volume is a further proof of the potentials that the Italian research regularly shows, thus contributing to the world-wide dimensions of the CL research.

