

# 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-ic.it](http://www.ai-ic.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 una realtà italiana 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

---

\* Dept. of Enterprise Engineering - 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. 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'Associazione Italiana di Linguistica Computazionale 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 sul risorse e tecnologie sviluppate nell'ambito della linguistica computazionale.

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

- una 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;
- un 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;
- un 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;
- una 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'Associazione Italiana di Linguistica Computazionale. 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 rilevanti 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 scien-

tifica 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 sottoporre 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 vettoriali, 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, una 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. Nella esaustiva analisi sperimentale discussa nell'articolo, si mostra come APAnt migliori le misure di metodi precedentemente pubblicati nel task di riconoscimento di antinomie.

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 della 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 Vittorio. 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.