

LINGUISTICA
COMPUTAZIONALE

VOLUME XVI-XVII

COMPUTATIONAL LINGUISTICS
IN PISA

*

LINGUISTICA COMPUTAZIONALE
A PISA

SPECIAL ISSUE

I

Editors:

Antonio Zampolli, Nicoletta Calzolari, Laura Cignoni



ISTITUTI EDITORIALI E POLIGRAFICI INTERNAZIONALI[®]

PISA - ROMA

LINGUISTICA
COMPUTAZIONALE

Editorial Board:

Andrea Bozzi, Nicoletta Calzolari,
Laura Cignoni, Eugenio Picchi, Vito Pirrelli,
Irina Prodanof, Giovanna Turrini, Antonio Zampolli

Linguistica Computazionale is published twice a year.

Subscriptions should be sent to the Publisher:

Istituti Editoriali e Poligrafici Internazionali[®], casella postale n.^o 1,
succursale n. 8, I, 56123 Pisa, Italy.

Pisa offices: Via Giosuè Carducci 60 · I 56010 Ghezzano · La Fontina
(Pisa)

Tel. +39 050878066 · Fax +39 050878732 · E-mail: iepi@iepi.it

Rome offices: Via Ruggero Bonghi 11/b (Colle Oppio) · I 00184 Roma

Tel +39 0670452494 · Fax +39 0670476605 · E-mail: iepi.roma@iepi.it

<http://www.iepi.it>

Subscription rates (one year): Euro 30,99 individuals (US\$ 60 abroad)
Euro 46,48 institutions (US\$ 90 abroad).

Subscriptions should be paid as follows:

- by cheque/international money order payable to: Istituti Editoriali e Poligrafici Internazionali[®];
- at our post office current account n. 13137567;
- by credit card (*American Express, Eurocard Mastercard, Visa*).

Direttore responsabile: Antonio Zampolli

Registrazione presso il tribunale di Pisa n. 1 dell'11/03/1981

The Editorial Board was assisted by Paola Orsolini who was responsible for the final copyediting work.

© Copyright 2003 by

Istituti Editoriali e Poligrafici Internazionali[®].

No reproduction of any part may take place without the written permission of Istituti Editoriali e Poligrafici Internazionali[®].

Our Publishing House guarantees for absolute discretion about personal information given by subscribers; on written request, this data can be modified or erased. The material, stored at our Publishing House, will only be used to inform subscribers about our new editorial enterprises (Italian Law 675/96).

ISSN 0392-6907

ISBN 88-8147-340-2

Contents

Foreword	xiii
Antonio Zampolli, Nicoletta Calzolari, Laura Cignoni	
Le principali attività dell'Istituto di Linguistica Computazionale	xvii
Antonio Zampolli	
Example-based automatic induction of semantic classes through entropic scores	1
Paolo Allegri, Simonetta Montemagni, Vito Pirrelli	
The <i>Diphilos</i> workstation: a computational system for digital philology	47
Andrea Bozzi, Maria Sofia Corradini	
Corpus-based lexicon building: an overview across projects, problems, approaches	79
Nicoletta Calzolari	
Acquisition of lexical information from a large textual Italian corpus	117
Nicoletta Calzolari, Remo Bindi	
Computer processing of a <i>Clopen</i> language system: old-church Slavonic	133
Giuseppe Camuglia, Monia Camuglia Ribarov, Kiril Ribarov	
Strumenti per l'accesso multilingue al contenuto dell'informazione	151
Amedeo Cappelli, Lorenzo Moretti, Maria Novella Catarsi, Clelia Caligaris	
OLISSIPO: strumento per l'estrazione automatica del vocabolario di base	183
Giuseppe Cappelli	

A corpus-based account of suffix productivity in Italian	201	Monica Monachini, Paola Orsolini, Eugenio Picchi, Sergio Rossi, Nicoletta Calzolari, Antonio Zampolli	
Francesca Carota, Irina Prodanof			
L'archivio elettronico delle opere di Carlo Emilio Gadda. Da redattori a fruitori di un data base testuale	221	Methods for standardization: the case of morphosyntax within the EAGLES project	423
Maria Luigia Ceccotti, Manuela Sassi		Monica Monachini, Nicoletta Calzolari	
At the interface of onomastics and phraseology. Multiword units as proper names, proper names as ‘common’ phrasal units	251	The syntactic-semantic Treebank of Italian. An Overview	461
Laura Cignoni, Stephen Coffey		Simonetta Montemagni, Francesco Barsotti, Marco Battista, Nicoletta Calzolari, Ornella Corazzari, Alessandro Lenci, Vito Pirrelli, Antonio Zampolli, Francesca Fanciulli, Maria Massetani, Remo Raffaelli, Roberto Basili, Maria Teresa Pazienza, Dario Saracino, Fabio Zanzotto, Nadia Mana, Fabio Pianesi, Rodolfo Delmonte	
Considerations emerging from a frequency study of multiword units in a corpus of contemporary written Italian	263	DBT-ALT: a system for storing and querying the data of the ‘Atlante Lessicale Toscano’	493
Laura Cignoni, Stephen Coffey		Simonetta Montemagni, Eugenio Picchi, Lisa Biagini	
Computational modelling of tutorial dialogue	285	BIBLOS: historical, philosophical and philological digital library of the Italian National Research Council	519
Giacomo Ferrari, Irina Prodanof		Gabriella Pardelli	
SIMPLE: plurilingual semantic lexicons for natural language processing	323	An Italian children’s corpus of spoken language	547
Alessandro Lenci, Nicoletta Calzolari, Antonio Zampolli		Laura Pecchia, Elisabetta Guazzini	
Chunk-it. An Italian shallow parser for robust syntactic annotation	353	Bilingual lexicons, parallel and comparable corpora: creating the basis for cross language information retrieval	573
Alessandro Lenci, Simonetta Montemagni, Vito Pirrelli		Carol Peters, Eugenio Picchi	
Per una storia dell’archivio testuale dell’Istituto di Linguistica Computazionale: dati e gestione del catalogo informatizzato	387	PiSystem: sistemi integrati per l’analisi testuale	597
Rita Marinelli		Eugenio Picchi	
The Italian PAROLE corpus: an overview	401		
Rita Marinelli, Lisa Biagini, Remo Bindi, Sara Goggi,			

Risorse monolingui e multilingui. Corpus bilingue italiano-arabo	629	A multi-level annotation meta-scheme for dialogue acts	925
Eugenio Picchi, Eva Sassolini, Ouafae Nahli, Sebastiana Cucurullo		Claudia Soria, Vito Pirrelli	
Syntagmatic and paradigmatic issues in computational morphology	679	From a children's dictionary to a hypermedia laboratory for language learning	953
Vito Pirrelli, Marco Battista		Giovanna Turrini, Laura Cignoni, Alessandro Paccosi	
Resources and tools: experiences in language engineering	703	Contributors	971
Irina Prodanof, Amedeo Cappelli, Lorenzo Moretti			
<i>ItalWordNet: building a large semantic database for the automatic treatment of Italian</i>	745		
Adriana Roventini, Antonietta Alonge, Francesca Bertagna, Nicoletta Calzolari, Jessica Cancila, Christian Girardi, Bernardo Magnini, Rita Marinelli, Manuela Speranza, Antonio Zampolli			
The PAROLE model and the Italian syntactic lexicon	793		
Nilda Ruimy, Ornella Corazzari, Elisabetta Gola, Antonietta Spanu, Nicoletta Calzolari, Antonio Zampolli			
A computational semantic lexicon of Italian: SIMPLE	821		
Nilda Ruimy, Monica Monachini, Elisabetta Gola, Nicoletta Calzolari, Maria Cristina Del Fiorentino, Marisa Olivieri, Sergio Rossi			
Morfsin and AyDA: two systems for analyzing modern and old Spanish	865		
Antonina Saba, Giuseppe Cappelli			
Letteratura, diritto e linguistica computazionale. Panorama delle collaborazioni Italia - Cuba	901		
Manuela Sassi, Yarina Amoroso			