Studi Interdisciplinari su
Traduzione Lingue e Culture

Studi Interdisciplinari su Traduzione, Lingue e Culture.

Collana a cura del Dipartimento di Studi Interdisciplinari su Traduzione, Lingue e Culture (SITLeC) dell'*Alma Mater Studiorum* – Università di Bologna, sede di Forlì.

La collana del SITLeC, fondata nel 2004, raccoglie le pubblicazioni prodotte nell'ambito dell'attività scientifica dei suoi afferenti e degli studiosi che operano negli stessi ambiti a livello nazionale e internazionale. Si propone come luogo editoriale di scambio e di dialogo interlinguistico e interculturale allo scopo di diffondere e rendere disponibili, a livello cartaceo e/o su supporto elettronico, i risultati della ricerca in diverse aree, come la linguistica teorica e applicata, la linguistica dei *corpora*, la terminologia, l'interpretazione, gli studi di genere, la critica letteraria, il teatro, gli studi culturali, gli studi sull'umorismo, privilegiando le dimensioni del tradurre, inteso come luogo di incontro e di scontro fra lingue e culture.

Tutte le pubblicazioni sono approvate dal Consiglio di Dipartimento, sentito il motivato parere di almeno due esperti qualificati esterni.

Il responsabile della Collana è il Direttore del SITLeC, affiancato da un comitato scientifico internazionale, articolato e variabile in relazione alle tematiche trattate.

WaCky!
Working Papers on the Web as Corpus

edited by
Marco Baroni and Silvia Bernardini

Gedit Edizioni

Collana Studi Interdisciplinari su Traduzione, Lingue e Culture
Direttore: Michele Prandi

Volume pubblicato con il contributo del Dipartimento di Studi Interdisciplinari su Traduzione Lingue e Culture (SITLeC) dell'*Alma Mater Studiorum* – Università di Bologna, sede di Forlì, Corso Diaz, 64 – 47100 Forlì.

ISBN 88-6027-004-9

First edition: September 2006

Gedit Edizioni
Via Irnerio 12/5
40126 Bologna
tel 051 4218740 fax 051 4210565
copertina: Avenida, Modena
stampa: Editografica, Rastignano (BO)

The volume was typeset in LaTeX by the editors and authors.

This collection of working papers puts together presentations at two Web as Corpus workshops (Forlì, January 14, 2005, Birmingham, July 13, 2005), and articles that were born out of discussions and collaborative experimentation among the WaCky community members. WaCky (for "**W**eb **a**s **C**orpus **k**ool **y**nitiative", in case you were wondering...) is a project started informally (i.e., with very little funding...) in 2003. It brings together linguists who think the World Wide Web is a great resource for their research, and that it would be even greater if it could be annotated and interrogated in a more linguist-friendly way. While we are aware that the task is an awesome one, we also believe that it is one worth putting some of our time and efforts into, and that interim results (e.g., the billion word, annotated, Web-derived corpora that have already seen the light for German and Italian) may equally provide very rich resources to study languages (on and off the Web). Through the publication of this collection of papers we hope to raise the interest of other researchers worldwide, who wish to contribute to this challenge.

For more information on WaCky or to participate in the initiative, please visit the WaCky wiki: `http://wacky.sslmit.unibo.it/`

We gratefully acknowledge the Fondazione Cassa dei Risparmi di Forlì for financial help in organizing the first Web as Corpus workshop. We also would like to thank the participants in the Web as Corpus workshops and in the online WaCky community – in particular, the contributors to this volume and Adam Kilgarriff – for very stimulating discussions.

Marco Baroni
Silvia Bernardini

List of Contributors

Marco Baroni SSLMIT, Università di Bologna, Corso della Repubblica 136, Forlì, 47100 Italy; baroni@sslmit.unibo.it

Silvia Bernardini SSLMIT, Università di Bologna, Corso della Repubblica 136, Forlì, 47100 Italy; silvia@sslmit.unibo.it

Sara Castagnoli SITLeC, Università di Bologna, Corso Diaz 64, Forlì, 47100 Italy; scastagnoli@sslmit.unibo.it

Massimiliano Ciaramita Yahoo! Research Barcelona, Ocata 1, 1st floor, 08003 Barcelona, Catalunya, Spain; massi@yahoo-inc.com

Thomas Emerson Gerson Lehrman Group, 2 Oliver Street, 7th Floor, Boston, MA 02109, USA; tree@glgroup.com

Stefan Evert Cognitive Science Institute, University of Osnabrück, Albrechtstr. 28, 49069 Osnabrück, Germany; stefan.evert@uos.de

Claudio Fantinuoli EURAC Research, Viale Druso 1, Bolzano, 39100 Italy; claudio.f@gmx.de

Rüdiger Gleim Bielefeld University, D-33615 Bielefeld, Germany; ruediger.gleim@uni-bielefeld.de

Alexander Mehler Bielefeld University, D-33615 Bielefeld, Germany; alexander.mehler@uni-bielefeld.de

John O'Neil Basis Technology, Inc., 150 CambridgePark Drive, Cambridge, MA 02140, USA; oneil@basistech.com

Serge Sharoff Centre for Translation Studies, School of Modern Languages and Cultures, University of Leeds, Leeds, LS2 9JT, UK; s.sharoff@leeds.ac.uk

Motoko Ueyama SSLMIT, Università di Bologna, Corso della Repubblica 136, Forlì, 47100 Italy; motoko@sslmit.unibo.it

Contents