

LANGUAGE RESOURCES AND EVALUATION

Volume 40 Nos. 3-4 2006

For detailed table of contents see pp. i-ii

Available  
online

[www.springerlink.com](http://www.springerlink.com)

Volume 40, nos. 3-4, 2006

ISSN 1574-020X


# Language Resources and Evaluation

Special Issue on:

Asian Language Processing:  
State-of-the Art Resources  
and Processing

Guest edited by:

Chu-Ren Huang  
Takenobu Tokunaga  
Sophia Y.M. Lee

 Springer

# LANGUAGE RESOURCES AND EVALUATION

## EDITORS-IN-CHIEF

Nancy Ide, *Department of Computer Science, Vassar College, Poughkeepsie, NY, USA*  
Nicoletta Calzolari, *Istituto di Linguistica Computazionale, CNR, Pisa, Italy*

## BOOK REVIEW EDITOR

Alessandro Lenci, *Università degli Studi di Pisa, Italy*

## ADVISORY BOARD

Bente Maegaard, *Center for Sprogteknologi, University of Copenhagen, Copenhagen, Denmark*  
Khalid Choukri, *ELRA, Paris, France*  
Joseph Mariani, *LIMS-CNRS, Orsay, France and Ministère Délégué à la Recherche, Paris, France*  
Jan Odijk, *Nuance Communications International Belgium, Merelbeke, Belgium and Utrecht University, Utrecht, The Netherlands*  
Christopher Cieri, *Linguistic Data Consortium, University of Pennsylvania, Philadelphia, USA*  
Junichi Tsujii, *University of Tokyo, Tokyo, Japan*

## EDITORIAL BOARD

Steven Bird, *University of Melbourne, Australia*; Paul Buitelaar, *DFKI GmbH, Germany*; Nick Campbell, *ATR Human Information Science Labs, Japan*; Key-Sun Choi, *KORTERMI/KAIST, Republic of Korea*; Kenneth Church, *Microsoft Research, USA*; Hamish Cunningham, *University of Sheffield, UK*; Tomas Erjavec, *Institute "Jozef Stefan", Ljubljana, Slovenia*; Christiane Fellbaum, *Princeton University, USA*; John Garofolo, *NIST, USA*; Dafydd Gibbon, *Universität Bielefeld, Germany*; Eduard Hovy, *University of Southern California, USA*; Chu-Ren Huang, *Academia Sinica, Taiwan*; Shuichi Itahashi, *National Institute of Advanced Industrial Science and Technology, Japan*; Adam Kilgarriff, *University of Sussex, UK*; Margaret King, *University of Geneva, Switzerland*; Gianni Lazzari, *ITC-irst, Italy*; Dekang Lin, *University of Alberta CA*; Rada Mihalcea, *University of North Texas, USA*; Asuncion Moreno, *Universitat Politècnica de Catalunya, Spain*; Martha Palmer, *University of Colorado at Boulder, USA*; Florence Reeder, *MITRE Corporation, USA*; Laurent Romary, *INRIA-Loria, France*; Florian Schiel, *Institut für Phonetik und Sprachliche Kommunikation der LMU München, Germany*; Gregor Thurmair, *Linguatex GmbH, Germany*; Takenobu Tokunaga, *Tokyo Institute of Technology, Japan*; Dan Tufis, *Romanian Academy of Sciences, Romania*; Janyce Wiebe, *University of Pittsburgh, USA*

*Language Resources and Evaluation* is published quarterly.

Subscriptions should be sent to **Springer Customer Service Journals, Haberstraße 7, 69126 Heidelberg, Germany**, or P.O. Box 2485, Secaucus, NJ 07094-2485, U.S.A., or to any subscription agent. Private subscriptions should be sent direct to the publishers. Changes of mailing address should be notified together with our latest label.

Periodicals postage paid at Rahway, N.J. USPS No. 746-490.

U.S. Mailing Agent: Mercury Airfreight International Ltd., 365 Blair Road, Avenel, NJ 07001, U.S.A. *Postmaster:* Please send all address corrections to: *Language Resources and Evaluation*, c/o: Mercury Airfreight International Ltd., 365 Blair Road, Avenel, NJ 07001, U.S.A.

For advertisement rates, prices of back volumes, and other information, apply to Springer, P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

**Photocopying.** *In the U.S.A.:* This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923 [info@copyright.com]. Authorizations to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Springer for users registered with the Copyright Clearance Center (CCC). For those organizations that have been granted a photocopy licence, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or for promotional purposes, for creating new collective works, or for resale. The CCC's services for users can be found on the internet at [www.copyright.com](http://www.copyright.com).

*In the rest of the world:* Permissions to photocopy must be obtained from the copyright owner. Please apply to Springer Rights & Permissions Department, P.O. Box 17, 3300 AA, Dordrecht, The Netherlands [permissions.dordrecht@springer.com].

# LANGUAGE RESOURCES AND EVALUATION

Volume 40 Nos. 3-4 2006

- CHU-REN HUANG, TAKENOBU TOKUNAGA and SOPHIA YAT MEI LEE / Asian language processing: current state-of-the-art 203-218
- ROSIE JONES, KEVIN BARTZ, PERO SUBASIC and BENJAMIN REY / Automatically generating related queries in Japanese 219-232
- MASAKI MURATA, QING MA, KIYOTAKA UCHIMOTO, TOSHIYUKI KANAMARU and HITOSHI ISAHARA / Japanese-to-English translations of tense, aspect, and modality using machine-learning methods and comparison with machine-translation systems on market 233-242
- CHIKARA HASHIMOTO, SATOSHI SATO and TAKEHITO UTSURO / Detecting Japanese idioms with a linguistically rich dictionary 243-252
- FRANCIS BOND, SANAE FUJITA and TAKA AKI TANAKA / The Hinoki syntactic and semantic treebank of Japanese 253-261
- TOMOHIRO OHNO, SHIGEKI MATSUBARA, HIDEKI KASHIOKA, TAKEHIKO MARUYAMA, HIDEKI TANAKA and YASUYOSHI INAGAKI / Dependency parsing of Japanese monologue using clause boundaries 263-279
- SO-YOUNG PARK, YOUNG-IN SONG and HAE-CHANG RIM / A segment-based annotation tool for Korean treebanks with minimal human intervention 281-289
- THI MINH HUYÈN NGUYÈN, LAURENT ROMARY, MATHIAS ROSSIGNOL and XUÂN LU'ONG VŨ / A lexicon for Vietnamese language processing 291-309
- SAJIB DASGUPTA and VINCENT NG / Unsupervised morphological parsing of Bengali 311-330
- PUSHPAK BHATTACHARYYA, DEBASRI CHAKRABARTI and VAIJAYANTHI M. SARMA / Complex predicates in Indian languages and wordnets 331-355

- JUN PENG BAO, CAROLINE LYON and PETER C. R. LANE / Copy detection in Chinese documents using Ferret 357-365
- JING-SHIN CHANG and WEI-LUN TENG / Mining atomic Chinese abbreviations with a probabilistic single character recovery model 367-374
- DONGHONG JI, YANXIANG HE and GUOZHENG XIAO / Word sense learning based on feature selection and MDL principle 375-393
- NIANWEN XUE / A Chinese semantic lexicon of senses and roles 395-403
- NIGEL COLLIER, AI KAWAZOE, LIHUA JIN, MIKA SHIGEMATSU, DINH DIEN, ROBERTO A. BARBERO, KOICHI TAKEUCHI and ASANEE KAWTRAKUL / A multilingual ontology for infectious disease surveillance: rationale, design and challenges 405-413

*Language Resources and Evaluation* is the first publication devoted to the acquisition, creation, annotation, and use of language resources, together with methods for evaluation of resources, technologies, and applications.

Language resources include language data and descriptions in machine readable form used to assist and augment language processing applications, such as written or spoken corpora and lexica, multimodal resources, grammars, terminology or domain specific databases and dictionaries, ontologies, multimedia databases, etc., as well as basic software tools for their acquisition, preparation, annotation, management, customization, and use.

Evaluation of language resources concerns assessing the state-of-the-art for a given technology, comparing different approaches to a given problem, assessing the availability of resources and technologies for a given application, benchmarking, and assessing system usability and user satisfaction.

*Printed on acid free paper*

**Photocopying.** *In the U.S.A.:* This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923 [info@copyright.com]. Authorizations to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Springer for users registered with the Copyright Clearance Center (CCC). For those organizations that have been granted a photocopy licence, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or for promotional purposes, for creating new collective works, or for resale. The CCC's services for users can be found on the internet at [www.copyright.com](http://www.copyright.com).

*In the rest of the world:* Permissions to photocopy must be obtained from the copyright owner. Please apply to Springer Rights & Permissions Department, P.O. Box 17, 3300 AA, Dordrecht, The Netherlands [permissions.dordrecht@springer.com].

*Language Resources and Evaluation* is a peer-reviewed journal.

*Language Resources and Evaluation* is published quarterly. Subscriptions should be sent to **Springer Customer Service Journals, Haberstraße 7, 69126 Heidelberg, Germany, or P.O. Box 2485, Secaucus, NJ 07094-2485 U.S.A.**, or to any subscription agent. Private subscriptions should be sent direct to the publishers. Changes of mailing address should be notified together with our latest label. Periodicals postage paid at Rahway, NJ. USPS No. 746-490.

U.S. Mailing Agent: Mercury Airfreight International Ltd., 365 Blair Road, Avenel, NJ 07001, U.S.A. *Postmaster:* Please send all address corrections to: *Computers and the Humanities*, c/o: Mercury Airfreight International Ltd., 365 Blair Road, Avenel, NJ 07001, U.S.A.

For advertisement rates, prices of back volumes, and other information, apply to Springer, P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Available  
online

[www.springerlink.com](http://www.springerlink.com)