

Computational Linguistics

Special Issue on Using Large Corpora: II

Articles

Using Register-Diversified Corpora for General Language Studies	Douglas Biber	219
From Grammar to Lexicon: Unsupervised Learning of Lexical Syntax	Michael R. Brent	243
The Mathematics of Statistical Machine Translation: Parameter Estimation	Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer	263
Building a Large Annotated Corpus of English: The Penn Treebank	Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz	313
Lexical Semantic Techniques for Corpus Analysis	James Pustejovsky, Sabine Bergler, and Peter Anick	331
Coping with Ambiguity and Unknown Words through Probabilistic Models	Ralph Weischedel, Marie Meteer, Richard Schwartz, Lance Ramshaw, and Jeff Palmucci	359

Book Reviews

An Introduction to Machine Translation W. John Hutchins and Harold L. Somers	<i>Reviewed by</i> Guðrún Magnúsdóttir	383
Meaning and Speech Acts Daniel Vanderveken	<i>Reviewed by</i> Amichai Kronfeld	385
Grammatical Competence and Parsing Performance Bradley L. Pritchett	<i>Reviewed by</i> Gerry Altmann	390
Principle-Based Parsing: Computation and Psycholinguistics Robert C. Berwick, Steven P. Abney, and Carol Tenny, editors	<i>Reviewed by</i> Geoffrey K. Pullum	393
Questions and Information Systems Thomas W. Lauer, Eileen Peacock, and Arthur C. Graesser, editors	<i>Reviewed by</i> John S. White	397

Books Received		401
Letters to the editor		405



Editor

James F. Allen
University of Rochester
Computer Science Department
Rochester, NY 14627, USA
(+1 716) 275-5288
acl@cs.rochester.edu

Managing Editor

Donald E. Walker
Bellcore, MRE 2A379
445 South Street, Box 1910
Morristown, NJ 07960-1910, USA
(+1 201) 829-4312
walker@bellcore.com

Associate Editor

Robert C. Berwick
Massachusetts Institute of Technology
Artificial Intelligence Laboratory
Cambridge, MA 02139, USA
(+1 617) 253-8918
berwick@ai.mit.edu

Book Review Editor

Graeme Hirst
University of Toronto
Computer Science Department
Toronto, CANADA M5S 1A4
(+1 416) 978-8747
gh@ai.toronto.edu

The FINITE STRING Editor

Ralph M. Weischedel
BBN Systems & Technologies
10 Moulton Street
Cambridge, MA 02138, USA
(+1 617) 873-3496
weischedel@bbn.com

Squibs and Discussions Co-Editors

James Pustejovsky
Brandeis University
Computer Science, Ford Hall
Waltham, MA 02254, USA
(+1 617) 736-2709
jamesp@cs.brandeis.edu

Robert J. P. Ingria
BBN Systems & Technologies
10 Moulton Street
Cambridge, MA 02138, USA
(+1 617) 873-3582
ingria@bbn.com

Editorial Board Members**1990–92:**

Kenneth Church, AT&T Bell Laboratories
Julia Hirschberg, AT&T Bell Laboratories
Robert Moore, SRI International
James Pustejovsky, Brandeis University
Stuart Shieber, Harvard University
Susan Warwick, University of Geneva

1991–93:

Eduard Hovy, USC/Information Sciences Institute
Paul Jacobs, General Electric
Robert Kasper, Ohio State University
Diane Litman, Columbia University
Graeme Ritchie, University of Edinburgh
Richard Sproat, AT&T Bell Laboratories
Oliviero Stock, Inst. for Scientific
& Technological Research

1992–94:

Jared Bernstein, SRI International
Edward J. Briscoe, University of Cambridge
Peter F. Brown, IBM/TJ Watson Research Center
Sandra Carberry, University of Delaware
Robert Dale, University of Edinburgh
Mary Dalrymple, Xerox-Palo Alto Research Center
Pierre Isabelle, CWARC Communications
Yuji Matsumoto, Kyoto University
John Nerbonne, German Research Center for AI
K. Vijay-Shanker, University of Delaware

Computational Linguistics

Special Issue on Using Large Corpora: II

Articles

Using Register-Diversified Corpora for General Language Studies	Douglas Biber	219
From Grammar to Lexicon: Unsupervised Learning of Lexical Syntax	Michael R. Brent	243
The Mathematics of Statistical Machine Translation: Parameter Estimation	Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer	263
Building a Large Annotated Corpus of English: The Penn Treebank	Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz	313
Lexical Semantic Techniques for Corpus Analysis	James Pustejovsky, Sabine Bergler, and Peter Anick	331
Coping with Ambiguity and Unknown Words through Probabilistic Models	Ralph Weischedel, Marie Meteer, Richard Schwartz, Lance Ramshaw, and Jeff Palmucci	359

Book Reviews

An Introduction to Machine Translation W. John Hutchins and Harold L. Somers	<i>Reviewed by</i> Guðrún Magnúsdóttir	383
Meaning and Speech Acts Daniel Vanderveken	<i>Reviewed by</i> Amichai Kronfeld	385
Grammatical Competence and Parsing Performance Bradley L. Pritchett	<i>Reviewed by</i> Gerry Altmann	390
Principle-Based Parsing: Computation and Psycholinguistics Robert C. Berwick, Steven P. Abney, and Carol Tenny, editors	<i>Reviewed by</i> Geoffrey K. Pullum	393
Questions and Information Systems Thomas W. Lauer, Eileen Peacock, and Arthur C. Graesser, editors	<i>Reviewed by</i> John S. White	397

Books Received 401

Letters to the editor 405

Copyright ©1993 by the Association for Computational Linguistics. The following code, 0891-2017/93, indicates the copyright owners' consent that a single copy of an article may be made for personal use. The consent is given, however, on the condition that the copier pay the stated per-copy fee of \$3.00 through the Copyright Clearance Center, Inc., 27 Congress Street, Salem, MA 01970, USA, for copying beyond that permitted by Sections 107 and 108 of the U.S. Copyright Law. This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. Such permission requests and other permission inquiries should be addressed to the Managing Editor at the address given on the inside front cover of this issue.

Computational Linguistics is abstracted and/or indexed in the following publications: *Annual Review of Information Science and Technology*, *Computer Abstracts*, *Computing Reviews*, *Current Contents: Social and Behavioral Sciences*, *Knowledge Engineering Review*, *Language and Language Behavior Abstracts*, *Linguistic Abstracts*, *Linguistic Bibliography*, *Science Abstracts* (INSPEC Section C), and *Social Sciences Citation Index*.

Volume 14 and subsequent volumes of *Computational Linguistics* are available on microfilm through University Microfilms International, 300 North Zeeb Road, Ann Arbor, MI 48106, USA.

Individual Subscriptions to *Computational Linguistics* are available only with membership in the Association for Computational Linguistics (ACL). Annual Membership Dues are as follows: *regular*, \$25.00; *student* — taking a full-time program of courses and not earning a regular income, \$15.00; *joint* — for couples receiving one subscription, \$30.00. A membership application together with an order form for back issues and conference proceedings is included at the end of this issue. Send membership applications and address changes to Dr. Donald E. Walker (ACL), Bellcore, MRE 2A379, 445 South Street, Box 1910, Morristown, NJ 07960-1910, USA. However, note that it is also possible to pay through banks in Switzerland and Japan.

Institutional Subscriptions are \$80.00. Outside the U.S. add \$14.00 postage and handling. Canadian subscribers also add 7% GST. Send institutional subscription orders and address changes to MIT Press Journals, 55 Hayward Street, Cambridge, MA 02142-1399, USA. To be honored free of charge, claims for missing copies must be made immediately upon receipt of the next published issue. Institutions should order back issues (before 1988) and all proceedings from the ACL at the address above.