



Cambridge Core

The new home of
Cambridge Journals
cambridge.org/core

Cambridge Core



CAMBRIDGE
UNIVERSITY PRESS



Computer Science

Books and Journals from
Cambridge University Press

Cambridge publish in a number of areas of
Computer Science including artificial intelligence, programming
languages, robotics and the application of mathematics to
computer science

For further details visit:
cambridge.org/core-computer-science

Cambridge
Core



CAMBRIDGE
UNIVERSITY PRESS

SUBSCRIPTIONS

Natural Language Engineering (ISSN 1351-3249; electronic 1469-8110) is published six times a year in January, March, May, July, September and November. Six parts form a volume. The subscription price (excluding VAT and sales tax) of volume 24 (2018) (which includes postage) is £387 net (US\$632 in the USA, Canada and Mexico) for institutions print and electronic, institutions electronic only £321/\$528; £46 net (US\$75 in the USA, Canada and Mexico) for individuals ordering direct from the publisher and certifying that the journal is for their personal use. Single parts are £71 (US\$119 in the USA, Canada and Mexico) plus postage.

Orders, which must be accompanied by payment, may be sent to a bookseller, subscription agent or direct to the publisher: Cambridge University Press, Journals Fulfillment Department, UPH, Shaftesbury Road, Cambridge CB2 8BS, UK; or in the USA, Canada and Mexico: Cambridge University Press, Journals Fulfillment Department, 1 Liberty Plaza, Floor 20, New York, NY 10006, USA. EU subscribers who are not registered for VAT should add VAT at their country's rate. VAT registered subscribers should provide their VAT registration number. Japanese prices for institutions are available from Kinokuniya Company Ltd, PO Box 55, Chitose, Tokyo 156, Japan.

COPYING

This journal is registered with the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, USA. Organizations in the USA who are also registered with C.C.C. may therefore copy material (beyond the limits permitted by sections 107 and 108 of U.S. Copyright law) subject to payment to C.C.C. of the per-copy fee of \$16.00. This consent does not extend to multiple copying for promotional or commercial purposes. Code 1351-3249/2015.

ISI Tear Sheet Service, 3501 Market Street, Philadelphia, PA 19104, USA, is authorised to supply single copies of separate articles for private use only.

Organisations authorised by the Copyright Licensing Agency may also copy material subject to the usual conditions.

For all other use, permission should be sought from Cambridge or from the American Branch of Cambridge University Press.

INSTRUCTIONS FOR CONTRIBUTORS

Instructions for Contributors will be found on the Journal's webpage at www.cambridge.org/nle. This journal and other journals are included in the Cambridge Journals Online service which can be found at journals.cambridge.org

© Cambridge University Press 2018

This journal issue has been printed on FSC-certified paper and cover board. FSC is an independent, non-governmental, non-for-profit organization established to promote the responsible management of the world's forests. Please see www.fsc.org for information.

NATURAL LANGUAGE ENGINEERING

Volume 24 Part 1 January 2018

CONTENTS

Editorial note <i>Prof Ruslan Mitkov, Executive Editor</i>	1
ARTICLES	
To use or not to use: Feature selection for sentiment analysis of highly imbalanced data <i>Sandra Kübler, Can Liu and Zeeshan Ali Sayyed</i>	3
A Semi-automatic and low-cost method to learn patterns for named entity recognition <i>M. Marrero and J. Urbano</i>	39
CCG supertagging with bidirectional long short-term memory networks <i>Rekia Kadari, Yu Zhang, Weinan Zhang and Ting Liu</i>	77
New treebank or repurposed? On the feasibility of cross-lingual parsing of Romance languages with Universal Dependencies <i>Marcos García, Carlos Gómez-Rodríguez and Miguel A. Alonso</i>	91
A survey of diacritic restoration in abjad and alphabet writing systems <i>Franklin Oḷádítípò Asahiah, Oḍétúnjí Àjàdí Oḍéjóbí and Emmanuel Rótímí Adágúnodò</i>	123
Emerging trends: A tribute to Charles Wayne <i>Kenneth Ward Church</i>	155
Book Review	
Formalizing Natural Languages: The NooJ Approach <i>Max Silberztein</i>	161

The cover illustration is computer generated from an engraving of the Tower of Babel in an eighteenth-century bible.

Cambridge Journals Online

For further information about this journal
please go to the journal website at:
cambridge.org/nle



MIX
Paper from
responsible sources
FSC® C007785

CAMBRIDGE
UNIVERSITY PRESS