Natural Language Engineering meets the needs of professionals and researchers working in all areas of automatic language processing, whether from the perspective of theoretical or corpus linguistics, translation, lexicography, computer science or engineering. Its aim is to bridge the gap between theoretical linguistic research and the implementation of practical applications with potential real-world use. As well as publishing original research articles on a broad range of topics - from text analysis, machine translation, information retrieval, speech processing and generation to integrated systems and multi-modal interfaces - it also publishes special issues on specific natural language processing methods, tasks or applications. The journal welcomes survey papers describing the state of the art of a specific topic. The Journal of Natural Language Engineering also publishes the popular Industry Watch and Emerging Trends columns as well as book reviews.

EDITORS
Ruslan Mitkov (Executive Editor), University of Wolverhampton, UK. Email: R.Mitkov@wlv.ac.uk
Branimir K. Bogataj, IBM Thomas J. Watson Research Center, USA. Email: bran@us.ibm.com

ASSOCIATE EDITORS
Diana Inkpen, University of Ottawa, Canada. Email: Diana.Inkpen@uottawa.ca and Sandra Köbler, University of Indiana, USA. Email: skuebler@indiana.edu

BOOK EDITOR
Diana M. Maynard, University of Sheffield, UK. Email: d.maynard@sheffield.ac.uk

INDUSTRY WATCH EDITOR
Robert Dale, Language Technology Group, Australia. Email: rdale@language-technology.com

EMERGING TRENDS EDITOR
Kenneth Church, Baidu, CA, USA. Email: kennethchurch@baidu.com

FOUNDING EDITORS
Roberto Gangemi, SenseGraph Ltd., UK.
John I. Tait, johntait.net Ltd., UK. Email: johntait.net

EDITORIAL BOARD
Kevin Bretonnel Cohen, University of Colorado, USA
Ondrej Bojar, Charles University, Czech Republic
Gregory Gefen, University of Florida for Human & Machine Cognition (HMC), USA
Tracy Holloway King, BS, Amazon, USA
Jing Jiang, Singapore Management University, Singapore
Sandu Köbler, University of Indiana, USA
Lori Lamel, LIMSI, France
Susobh Musaardibar, NAAM Institute, Nepal
Alessandro Moschitti, Amazon, USA
Prasen Naray, Qatar Computing Research Institute, HBKU, Qatar
Roberto Navigli, Sapienza University of Rome, Italy
Kemal Oflazer, Carnegie Mellon University, Qatar
Constantin Oraian, University of Surrey, UK
Sebastian Pado, University of Stuttgart, Germany
Simone Paolo Ponzoletto, University of Mannheim, Germany
Nasenide Seminar, CEA LETI, France
Khalil Sima’an, University of Amsterdam, The Netherlands
Richard Sproat, Google, USA
Tomek Strzalkowski, University at Albany, USA
Maite Taboada, Simon Fraser University, Canada
Hwee Tou Ng, National University of Singapore, Singapore
Allison Ureña López, University of Jaen, Spain
Aline Villavicencio, Federal University of Rio Grande do Sul, Brazil and University of Sheffield, UK
Xiaojun Wan, Peking University, China
Bonnie Webber, University of Edinburgh, UK
Min Zhang, Soochow University, China

EDITORIAL ASSISTANTS
Sandra Elfiky, University of Wolverhampton, UK
April Harper, University of Wolverhampton, UK
Sara Mozaffari, University of Wolverhampton, UK
Maria Kunilkovskaya, University of Wolverhampton, UK
Rocio Caro Quintana University of Wolverhampton
Bransilava Sandrich University of Wolverhampton and University of Belgrade

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 26 (2020) (which includes postage) is $428 net (US$698 in the USA, Canada and Mexico) for institutions print and electronic, institutions electronic only $354/$583; $52 net (US$83 in the USA, Canada and Mexico) for individual subscribers ordering direct from the publisher and certifying that the journal is for their personal use. Single parts are $79 (US$132 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, P.O. Box 35, Chatsworth, London NW10 6XU, UK. Science journals are shipped from the printer, and are sent to agents on receipt of payment. We do not ship to individuals ordering direct from the publisher and certifying that the journal is for their personal use. Single parts are $79 (US$132 in the USA, Canada and Mexico) plus postage.

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

© Cambridge University Press 2020

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

**INTRODUCTION**
Preface 593
Marcos Zampieri and Preslav Nakov

**SURVEY PAPER**
Natural language processing for similar languages, varieties, and dialects: A survey 595
Marcos Zampieri, Preslav Nakov and Yves Scherrer

**ARTICLES**
Classification of regional and genre varieties of Chinese: A correspondence analysis approach based on comparable balanced corpora 613
Renkui Hou and Chu-Ren Huang

Fine-grained analysis of language varieties and demographics 641
Francisco Rangel, Paolo Rosso, Wajdi Zaghouani and Anis Charfi

Compression versus traditional machine learning classifiers to detect code-switching in varieties and dialects: Arabic as a case study 663
Taghreed Tarmoan, William Toshar, Eric Atwell and Mohammed Anmar Alsaika

Effective multi-dialectal arabic POS tagging 677
Kareem Darwish, Mohammed Altra, Handuy Mubarak, Younes Samih, Ahmed Abdelali, Lluís Márquez, Mohamed Eldesouki and Laura Kallmeyer

Spoken Arabic dialect recognition using X-vectors 691
Abualsoud Hanani and Rabee Naser