

# **Workshop on Multiword Expressions: from Theory to Applications**

**(MWE 2010)**

**At the 23rd International Conference on Computational Linguistics 2010**

**Beijing, China  
48 August 2010**

**ISBN: 978-1-61782-343-5**

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2010) by the Chinese Information Processing Society of China  
All rights reserved.

Printed by Curran Associates, Inc. (2011)

For permission requests, please contact the Chinese Information Processing Society of China  
at the address below.

Chinese Information Processing Society of China  
No. 4 Zhongguancun South 4th St.  
Haidian District Beijing  
China 100190

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2634  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Being Theoretical is Being Practical: Multiword Units and Terminological Structure Revitalised</i> Kyo Kageura .....	1
<i>Computational Lexicography of Multi-Word Units. How Efficient Can It Be?</i> Filip Gralinski, Agata Savary, Monika Czerepowicka and Filip Makowiecki .....	2
<i>Construction of Chinese Idiom Knowledge-base and Its Applications</i> Lei Wang and Shuwen Yu .....	11
<i>Automatic Extraction of Arabic Multiword Expressions</i> Mohammed Attia, Antonio Toral, Lamia Tounsi, Pavel Pecina and Josef van Genabith .....	19
<i>Sentence Analysis and Collocation Identification</i> Eric Wehrli, Violeta Seretan and Luka Nerima .....	28
<i>Automatic Extraction of Complex Predicates in Bengali</i> Dipankar Das, Santanu Pal, Tapabrata Mondal, Tanmoy Chakraborty and Sivaji Bandyopadhyay	37
<i>Handling Named Entities and Compound Verbs in Phrase-Based Statistical Machine Translation</i> Santanu Pal, Sudip Kumar Naskar, Pavel Pecina, Sivaji Bandyopadhyay and Andy Way .....	46
<i>Application of the Tightness Continuum Measure to Chinese Information Retrieval</i> Ying Xu, Randy Goebel, Christoph Ringlstetter and Grzegorz Kondrak .....	55
<i>Standardizing Complex Functional Expressions in Japanese Predicates: Applying Theoretically-Based Paraphrasing Rules</i> Tomoko Izumi, Kenji Imamura, Genichiro Kikui and Satoshi Sato .....	64
<i>Identification of Reduplication in Bengali Corpus and their Semantic Analysis: A Rule Based Approach</i> Tanmoy Chakraborty and Sivaji Bandyopadhyay .....	73
<i>Contrastive Filtering of Domain-Specific Multi-Word Terms from Different Types of Corpora</i> Francesca Bonin, Felice Dell’Orletta, Giulia Venturi and Simonetta Montemagni .....	77
<i>A Hybrid Approach for Functional Expression Identification in a Japanese Reading Assistant</i> Gregory Hazelbeck and Hiroaki Saito .....	81
<i>An Efficient, Generic Approach to Extracting Multi-Word Expressions from Dependency Trees</i> Scott Martens and Vincent Vandeghinste .....	85
<i>Multiword Expressions as Discourse Relation Markers (DRMs)</i> Aravind Joshi .....	89