Workshop on Multiword Expressions: from Parsing and Generation to the Real World 2011

(MWE 2011)

Held at ACL HLT 2011

Portland, Oregon, USA
23 June 2011

ISBN: 978-1-61839-253-4

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© (2011) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2011)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

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
Web: www.proceedings.com

Table of Contents

MWEs and Topic Modelling: Enhancing Machine Learning with Linguistics
Invited talk by Timothy Baldwin 1

Automatic Extraction of NV Expressions in Basque: Basic Issues on Cooccurrence Techniques
Antton Gurrutxaga and Iñaki Alegria 2

Semantic Clustering: an Attempt to Identify Multiword Expressions in Bengali
Tanmoy Chakraborty, Dipankar Das and Sivaji Bandyopadhyay 8

Decreasing Lexical Data Sparsity in Statistical Syntactic Parsing - Experiments with Named Entities
Deirdre Hogan, Jennifer Foster and Josef van Genabith 14

Detecting Multi-Word Expressions Improves Word Sense Disambiguation
Mark Finlayson and Nidhi Kulkarni 20

Tree-Rewriting Models of Multi-Word Expressions
William Schuler and Aravind Joshi 25

Learning English Light Verb Constructions: Contextual or Statistical
Yuancheng Tu and Dan Roth 31

Two Types of Korean Light Verb Constructions in a Typed Feature Structure Grammar
Juwon Lee 40

MWU-Aware Part-of-Speech Tagging with a CRF Model and Lexical Resources
Matthieu Constant and Anthony Sigogne 49

The Web is not a PERSON, Berners-Lee is not an ORGANIZATION, and African-Americans are not LOCATIONS: An Analysis of the Performance of Named-Entity Recognition
Robert Krovetz, Paul Deane and Nitin Madnani 57

A Machine Learning Approach to Relational Noun Mining in German
Berthold Crysmann 65

Identifying and Analyzing Brazilian Portuguese Complex Predicates
Magali Sanches Duran, Carlos Ramisch, Sandra Maria Aluísio and Aline Villavicencio 74

An N-gram Frequency Database Reference to Handle MWE Extraction in NLP Applications
Patrick Watrin and Thomas François 83

Extracting Transfer Rules for Multiword Expressions from Parallel Corpora
Petter Haugereid and Francis Bond 92

Identification and Treatment of Multiword Expressions Applied to Information Retrieval
Otavio Acosta, Aline Villavicencio and Viviane Moreira 101

Stepwise Mining of Multi-Word Expressions in Hindi
Rai Mahesh Sinha 110

Detecting Noun Compounds and Light Verb Aonstructions: a Contrastive Study
Veronika Vincze, István Nagy T. and Gábor Berend 116

jMWE: A Java Toolkit for Detecting Multi-Word Expressions
Nidhi Kulkarni and Mark Finlayson 122

FipsCoView: On-line Visualisation of Collocations Extracted from Multilingual Parallel Corpora
Violeta Seretan and Eric Wehrli 125

The StringNet Lexico-Grammatical Knowledgebase and its Applications
David Wible and Nai-Lung Tsao 128

The Ngram Statistics Package (Text::NSP) : A Flexible Tool for Identifying Ngrams, Collocations, and Word Associations
Ted Pedersen, Satanjeev Banerjee, Bridget McInnes, Saiyam Kohli, Mahesh Joshi and Ying Liu 131

Fast and Flexible MWE Candidate Generation with the mwetoolkit
Vitor De Araujo, Carlos Ramisch and Aline Villavicencio 134

How Many Multiword Expressions do People Know?
Invited talk by Kenneth Church 137