



P A R S  M E



EUROPHRAS 2015

29 June - 2 July 2015, Malaga, Spain

Workshop Proceedings  
MULTI-WORD UNITS IN MACHINE  
TRANSLATION AND  
TRANSLATION TECHNOLOGIES  
MUMTTT2015

1-2 July 2015

**Editors:** Gloria Corpas Pastor, Johanna Monti, Violeta Seretan, Ruslan Mitkov



P A R S  M E



ISBN 9782970073697



2016. Editions Tradulex, Geneva

© The editors.

ISBN: 978-2-9700736-9-7

Distribution without the authorisation from the editors is not allowed.

This document is downloadable from [www.tradulex.com](http://www.tradulex.com).



## ORGANISING COMMITTEE

### Workshop Chairs

Gloria Corpas Pastor (University of Malaga, Spain)

Johanna Monti (Università degli Studi di Sassari, Italy)

Violeta Seretan (Université de Genève, Switzerland)

Ruslan Mitkov (University of Wolverhampton, United Kingdom)

### Advisory Board

Dmitrij O. Dobrovolskij (Russian Academy of Sciences, Russia)

Carlos Ramisch (Aix-Marseille University, France)

Michael Rosner (University of Malta, Malta)

Agata Savary (Université François Rabelais Tours, France)

Kathrin Steyer (Institut für Deutsche Sprache, Germany)

### Organisers

Rosario Bautista Zambrana

Cristina Castillo Rodríguez

Hernani Costa

Isabel Durán Muñoz

Jorge Leiva Rojo

Gema Lobillo Mora

Pablo Pérez Pérez Míriam

Seghiri Domínguez

Cristina Toledo Báez

Míriam Urbano Mendaña

Anna Zaretskaya

### Secretary

Míriam Buendía Castro

Rut Gutiérrez Florido



## Programme Committee

Iñaki Alegria (Universidad del País Vasco, Spain)  
Giuseppe Attardi (Università di Pisa, Italy)  
Doug Arnold (University of Essex, United Kingdom)  
António Branco (Universidade de Lisboa, Portugal)  
Paul Buitelaar (National University of Ireland, Galway)  
František Čermák (Univerzita Karlova v Praze, Czech Republic)  
Jean-Pierre Colson (Université Catholique de Louvain, Belgium)  
Matthieu Constant (Université Paris-Est, France)  
Gaël Dias (Université de Caen Basse-Normandie, France)  
Mike Dillinger (Association for Machine Translation in the Americas)  
Dmitrij O. Dobrovolskij (Russian Academy of Sciences, Russia)  
Peter Ďurčo (Univerzita sv. Cyrila a Metoda v Trnave, Slovak Republic)  
Marcello Federico (Fondazione Bruno Kessler, Italy)  
Sabine Fiedler (Universität Leipzig, Germany)  
Natalia Filatkina (Universität Trier, Germany)  
Thierry Fontenelle (Translation Centre for the Bodies of the European Union, Luxembourg)  
Corina Forăscu (Al.I. Cuza University of Iasi, Romania)  
Thomas François (Université Catholique de Louvain, Belgium)  
Ulrich Heid (Universität Hildesheim, Germany)  
Kyo Kageura (University of Tokyo, Japan)  
Cvetana Krstev (University of Belgrade, Serbia)  
Koenraad Kuiper (University of Canterbury, New Zealand)  
Alon Lavie (Carnegie Mellon University, USA)  
Malvina Nissim (Università di Bologna, Italy)  
Preslav Nakov (Qatar Computing Research Institute, Qatar Foundation, Qatar)  
Michael Oakes (University of Wolverhampton, United Kingdom)  
Adriana Orlandi (Università degli studi di Modena e Reggio Emilia, Italy)  
Yannick Parmentier (Université d'Orléans, France)  
Pavel Pecina (Univerzita Karlova v Praze, Czech Republic)  
Carlos Ramisch (Université de Grenoble, France)  
Victoria Rosén (Universitetet i Bergen, Norway)  
Michael Rosner (University of Malta)



Manfred Sailer (Goethe Universität, Germany)  
Tanja Samardžić (Universität Zurich, Switzerland)  
Agata Savary (Université François Rabelais Tours, France)  
Gerold Schneider (Universität Zurich, Switzerland)  
Gilles Sérasset (Université de Grenoble, France)  
Max Silberztein (Universié de Franche-Comté, France)  
Kiril Simov (Bulgarian Academy of Sciences, Bulgaria)  
Kathrin Steyer (Institut für Deutsche Sprache, Germany)  
Joanna Szerszunowicz (University of Bialystok, Poland)  
Marko Tadić (University of Zagreb, Croatia)  
Amalia Todirascu (Université de Strasbourg, France)  
Beata Trawinski (Institut für Deutsche Sprache Mannheim, Germany)  
Dan Tufiş (Romanian Academy, Romania)  
Agnès Tutin (Université de Grenoble, France)  
Lonneke van der Plas (Universität Stuttgart, Germany)  
Veronika Vincze (University of Szeged, Hungary)  
Martin Volk (Universität Zurich, Switzerland)  
Eric Wehrli (Université de Genève, Switzerland)  
Michael Zock (Aix-Marseille Université, France)



## MUMTTT2015 Academic Programme

**Wednesday, 1 July 2015: 15.30-17:00**

### **Session 1**

#### **Automatic extraction of multilingual MWU resources**

Chaired by Johanna Monti

*Bilingual Term Alignment for Language Service in Comparable Corpus*  
Zhiyan Zhao; Xuri Tang

*Building a Lexical Bundle Resource for MT*  
Natalia Grabar; Marie-Aude Lefer

*Assessing WordNet for Bilingual Compound Dictionary Extraction*  
Carla Parra Escartín; Héctor Martínez Alonso

**Wednesday, 1 July 2015: 17:20-18:50**

### **Session 2**

#### **Identification, acquisition and evaluation of multi-word terms**

Chaired by Violeta Seretan

*Multiword Units Translation Evaluation in Machine Translation: another pain in the neck?*  
Johanna Monti; Amalia Todirascu

*Identifying Multi-Word Expressions from Parallel Corpora with String-Kernels and Alignment Methods*  
Federico Sangati; Andreas van Cranenburgh; Mihael Arcan; Johanna Monti; Marco Turchi

*Parallel Sentences Fragmentation*  
Sergey Potemkin; Galina Kedrova



**Thursday, 2 July 2015: 9:00-10:00**

**Session 3**

**Multilingualism and MWU processing. MWUs and word alignment techniques**

Chaired by Ismail El Maarouf

*Statistical Measures to characterise MWUs involving 'mordre' in French or 'bite' in English*

Ismail El Maarouf; Michael Oakes

*Aligning Verb+Noun Collocations to Improve a French - Romanian FSMT System*

Amalia Todirascu; Mirabela Navlea

**Thursday, 2 July 2015: 12:20-13:30**

**Session 4**

**Learning semantic information about MWUs from monolingual, parallel or comparable corpora**

Chaired by Eric Wehrli

*Aranea: Comparable Gigaword Web Corpora*

Vladimír Benko; Peter Ďurčo

*In-depth Study of the Phraseological Units in Islamic and Christian Religions in Samples (corpora) of Religious Texts*

Madian Souliman; Ali Ahmad

**Thursday, 2 July 2015: 15:00-16:30**

**Session 5**

**MWUs in Machine Translation**

Chaired by Amalia Todirascu

*Multiword Expressions in Machine Translation: The case of German compounds*

Maria Ivanova; Eric Wehrli; Luka Nerima

*Adding Multi-Word Expressions More Than Once in Machine Translation*

Liling Tan; Josef van Genabith

*Multi-word Expressions in user-generated Content: How many and how well translated? Evidence from a Post-editing Experiment*

Violeta Seretan



**Thursday, 2 July 2015: 17:00-18:30**

**Session 6**

**Lexical, syntactic, semantic and translational aspects in MWU representation**

Chaired by Carla Parra Escartín

*Transformation and MWU in Quechua*

Maximiliano Durán

*Populating a Lexicon with Multiword Expressions in view of Machine Translation*

Voula Gioli

*MWEs: Light verb constructions vs Frozen constructions in Modern Greek and French*

Angeliki Fotopoulou, Voula Giouli

**Poster session**

*Challenges on the Translation of Collocations*

Angela Costa

*Chunking-based detection of English nominal compounds*

Gábor Csernyi

*Integrating Multi-Word Expressions in Statistical Machine Translation*

Zied Elloumi, Laurent Besacier, Olivier Kraif

*Analysis of Multiword Expression translation errors in Statistical Machine Translation*

Natalia Klyueva, Jeevanthi Liyanapathira





## Table of Contents

|   |    |
|---|----|
| Preface by the Workshop Chairs .....  | 12 |
| <b>Keynote talk</b>   |    |
| <i>Corpus-driven Description of Multi-word Patterns</i><br>Kathrin Steyer .....   | 13 |
| <b>Session 1 - Automatic extraction of multilingual MWU resources</b>   |    |
| <i>Assessing WordNet for Bilingual Compound Dictionary Extraction</i><br>Carla Parra Escartín; Héctor Martínez Alonso .....                                     | 19 |
| <b>Session 2 - Identification, acquisition and evaluation of multi-word terms</b>   |    |
| <i>Multiword Units Translation Evaluation in Machine Translation: Another Pain in the Neck?</i><br>Johanna Monti; Amalia Todirascu .....                        | 25 |
| <b>Session 3 - Multilingualism and MWU processing. MWUs and word alignment techniques</b>   |    |
| <i>Statistical Measures to Characterise MWE Involving 'mordre' in French or 'bite' in English</i><br>Ismail El Maarouf; Michael Oakes .....                     | 30 |
| <i>Aligning Verb+Noun Collocations to Improve a French - Romanian FSMT System</i><br>Amalia Todirascu; Mirabela Navlea .....                                    | 37 |
| <b>Session 5 - MWUs in Machine Translation</b>  |    |
| <i>Multiword Expressions in Machine Translation: The Case of German Compounds</i><br>Maria Ivanova; Eric Wehrli; Luka Nerima .....                              | 43 |
| <i>Multi-word Expressions in User-generated Content: How Many and How Well Translated?<br/>Evidence from a Post-editing Experiment</i><br>Violeta Seretan ..... | 50 |



## **Session 6 - Lexical, syntactic, semantic and translational aspects in MWU representation**

|   |    |
|---|----|
| <i>Transformation and Multi-word Units in Quechua</i><br>Maximiliano Durán .....  | 55 |
| <i>Populating a Lexicon with Multi-word Expressions in view of Machine Translation</i><br>Voula Giouli .....                                    | 61 |
| <i>MWEs: Support/light Verb Constructions vs Frozen Constructions in Modern Greek and French</i><br>Angeliki Fotopoulou, Voula Giouli .....     | 68 |
| <b>Poster session</b>   |    |
| <i>Challenges on the Automatic Translation of Collocations</i><br>Angela Costa .....  | 73 |
| <i>Chunking-based Detection of English Nominal Compounds</i><br>Gábor Csernyi .....   | 78 |
| <i>Integrating Multi-word Expressions in Statistical Machine Translation</i><br>Zied Elloumi, Laurent Besacier, Olivier Kraif .....             | 83 |
| <i>Analysis of Multiword Expression Translation Errors in Statistical Machine Translation</i><br>Natalia Klyueva, Jeevanthi Liyanapathira ..... | 87 |