Lecture Notes in Artificial Intelligence 11755

Subseries of Lecture Notes in Computer Science

Series Editors

Randy Goebel
University of Alberta, Edmonton, Canada
Yuzuru Tanaka
Hokkaido University, Sapporo, Japan
Wolfgang Wahlster
DFKI and Saarland University, Saarbrücken, Germany

Founding Editor

Jörg Siekmann
DFKI and Saarland University, Saarbrücken, Germany

More information about this series at http://www.springer.com/series/1244

Gloria Corpas Pastor · Ruslan Mitkov (Eds.)

Computational and Corpus-Based Phraseology

Third International Conference, Europhras 2019
Malaga, Spain, September 25–27, 2019
Proceedings

Springer

Editors
Gloria Corpas Pastor
University of Malaga
Malaga, Spain

Ruslan Mitkov
University of Wolverhampton
Wolverhampton, UK

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-030-30134-7 ISBN 978-3-030-30135-4 (eBook)
https://doi.org/10.1007/978-3-030-30135-4

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer Nature Switzerland AG 2019
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Computational and Corpus-Based Phraseology

Language is phraseological and phraseology is the discipline which studies phraseological units (PUs) and their related concepts referred to (and regarded largely synonymous) by scholars as multiword units, multiword expressions (MWEs), fixed expressions, set expressions, formulaic language, phrasemes, idiomatic expressions, idioms, collocations, and/or polylexical expressions. PUs or MWEs are ubiquitous and pervasive in language. They are a fundamental linguistic concept which is central to a wide range of natural language processing and applied linguistics applications, including, but not limited to, phraseology, terminology, translation, language learning, teaching and assessment, and lexicography. Jackendoff (1997) observes that the number of MWEs in a speaker's lexicon is of the same order of magnitude as the number of single words. Biber et al. (1999) argue that they constitute up to 45% of spoken English and up to 21% of academic prose in English. Sag et al. (2002) comment that they are overwhelmingly present in terminology and 41% of the entries in WordNet 1.7 are reported to be MWEs.

It is not only in the computational treatment of natural language that PUs play a crucial role. Terms are often MWEs (and not single words), which makes them highly relevant to terminology. Translation and interpreting are two other fields where phraseology plays an important role, as finding correct translation equivalents of PUs is a pivotal step in the translation process. Given their pervasive nature, PUs are absolutely central to the work carried out by lexicographers, who analyse and describe both single words and PUs. Last but not least, PUs are vital not only for language learning, teaching, and assessment, but also for more theoretical linguistic areas such as pragmatics, cognitive linguistics, and construction grammars. All the aforementioned areas are today aided by (and often driven by) corpora, which makes PUs particularly relevant for corpus linguists. Finally, PUs provide an excellent basis for inter- and multidisciplinary studies, fostering fruitful collaborations between researchers across different disciplines. While there are some exceptions (e.g. Corpas Pastor and Colson 2019/forthcoming; Granger and Meunier 2008; Mitkov et al. 2018), such collaborations are still generally largely unexplored.

This volume features a selection of regular papers accepted after rigorous reviewing and presented at the international conference "Computational and Corpus-based Phraseology" (Europhras 2019). The conference was jointly organised by the European Association for Phraseology (EUROPHRAS), the University of Malaga (Research Group in Lexicography and Translation), Spain, the University of Wolverhampton (Research Group in Computational Linguistics), United Kingdom, and the Association for Computational Linguistics, Bulgaria, and was sponsored by EUROPHRAS, the Sketch Engine, the University of Malaga, and the University of Wolverhampton. Europhras 2019 provided the perfect opportunity for researchers to present their work,

fostering interaction and interdisciplinary collaboration. The papers in this volume cover a number of topics including general corpus-based approaches to phraseology, phraseology in translation and cross-linguistic studies, phraseology in language teaching and learning, phraseology in specialised languages, phraseology in lexicography, cognitive approaches to phraseology, the computational treatment of multiword expressions, and the development, annotation and exploitation of corpora for phraseological studies. We have deliberately not grouped the contributions around specific topics – as many of them cut across more than one of the above listed topics – but have instead opted for presenting the papers alphabetically according to the (first) family name of the (first) author.

Every submission to the conference was evaluated by three reviewers. The reviewers were either members of the Programme Committee, consisting of 56 scholars from 24 different countries, or one of the 16 additional reviewers from 3 countries, who were recommended by the Programme Committee. The conference contributions were authored by a total of 150 scholars from 27 different countries. These figures attest to the truly international nature of Europhras 2019.

We would like to thank all colleagues who made this wonderfully interdisciplinary and international event possible. We would like to start by thanking all colleagues who submitted papers to Europhras 2019 and travelled to Malaga to attend the event. We are grateful to all members of the Programme Committee and the additional reviewers for carefully examining all submissions and providing substantial feedback on all papers, helping the authors of accepted papers to improve and polish the final versions of their papers. A special thanks goes to the invited speakers, namely the keynote speakers of the main conference (Sylviane Granger, Miloš Jakubíček, Natalie Kübler, Kathrin Steyer, and Aline Villavicencio), the invited speakers of the accompanying workshop on Multiword Units in Machine Translation and Translation Technology (Jean-Pierre Colson and Aline Villavicencio), and the tutorial speakers from Lexical Computing. Words of gratitude go to our sponsors: EUROPHRAS, the Sketch Engine, the University of Malaga, and the University of Wolverhampton.

Last but not least, we would like to use this paragraph to acknowledge the members of both the Organising Committee and Programme Committee, who worked very hard during the last 12 months and whose dedication and efforts made the organisation of this event possible. All those who have contributed are listed (in alphabetical order) on the following pages. We would like to single out several colleagues for competently carrying out numerous organisational and reviewing-related tasks and being ready to step in and support the organisation of the conference both on weekdays and during the weekends – in fact, whenever needed. Our big ‘thank you’ goes out to Rocío Caro Quintana, Maria Kunilovskaya, Francisco Javier Lima Florido, Desiré Martos García, Tharindu Ranasinghe Hettiarachchige, and Shiva Taslimipoor.

September 2019

Gloria Corpas Pastor
Ruslan Mitkov

References

Biber, D., Finegan, E., Johansson, S., Conrad, S. and Leech, G. 1999. *Longman Grammar of Spoken and Written English*. Longman, Harlow.

Corpas Pastor, G. and J.P. Colson (Eds). 2019/forthcoming. *Computational Phraseology*. Amsterdam and Philadelphia: John Benjamins.

Granger, S. and Meunier, F. 2008. Disentangling the Phraseological Web. In Granger, S., & Meunier, F. (Eds.), *Phraseology. An interdisciplinary perspective*. Amsterdam: John Benjamins publishers.

Jackendoff, R. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Mitkov, R., Monti, J., Corpas Pastor G. and Seretan, V. (Eds). 2018. *Multiword Units in Machine Translation and Translation Technology*. John Benjamins.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A. and Flickinger, D. 2002. 'Multiword Expressions: A Pain in the Neck for NLP'. In *Proceedings of the third international conference on intelligent text processing and computational linguistics (CICLING 2002)* (pp. 1−15). Mexico City, Mexico.

Organisation

Europhras 2019 was jointly organised by the European Association for Phraseology (Europhras), the University of Malaga (Research Group in Lexicography and Translation), Spain, the University of Wolverhampton (Research Group in Computational Linguistics), United Kingdom, and the Association for Computational Linguistics, Bulgaria.

Conference Chairs

Gloria Corpas Pastor	University of Malaga, Spain
Ruslan Mitkov	University of Wolverhampton, UK

Programme Committee

Mariangela Albano	University Dokuz Eylül of Izmir, Turkey
Verginica Barbu Mititelu	Romanian Academy Research Institute for Artificial Intelligence, Romania
Farouk Bouhadiba	University of Oran 2, Algeria
Nicoletta Calzolari	Institute for Computational Linguistics, Italy
María Luisa Carrió Pastor	Polytechnic University of Valencia, Spain
Sheila Castilho	Dublin City University, Ireland
Cristina Castillo Rodríguez	University of Malaga, Spain
Ken Church	Baidu, China
Jean-Pierre Colson	Université Catholique de Louvain, Belgium
Anna Čermáková	Charles University, Czech Republic
María Sagrario del Río Zamudio	University of Udine, Italy
Dmitrij Dobrovolskij	Russian Language Institute, Russia
Peter Ďurčo	University of St. Cyril and Methodius, North Macedonia
Jesse Egbert	Northern Arizona University, USA
Natalia Filatkina	University of Trier, Germany
Thierry Fontenelle	Translation Centre for the Bodies of the European Union, Luxembourg
José Enrique Gargallo	University of Barcelona, Spain
Sylviane Granger	Université Catholique de Louvain, Belgium
Kleanthes Grohmann	University of Cyprus, Cyprus
Miloš Jakubíček	Lexical Computing, Czech Republic
Simon Krek	University of Ljubljana, Slovenia
Natalie Kübler	Paris Diderot University, France
Alessandro Lenci	University of Pisa, Italy

Elvira Manero	University of Murcia, Spain
Carmen Mellado Blanco	University of Santiago de Compostela, Spain
Flor Mena Martínez	University of Murcia, Spain
Pedro Mogorrón Huerta	University of Alicante, Spain
Johanna Monti	L'Orientale University of Naples, Italy
Sara Može	University of Wolverhampton, UK
Michael Oakes	University of Wolverhampton, UK
Inés Olza	University of Navarra, Spain
Petya Osenova	Sofia University, Bulgaria
Stéphane Patin	Paris Diderot University, France
Alain Polguère	University of Lorraine, France
Encarnación Postigo Pinazo	University of Malaga, Spain
Carlos Ramisch	Laboratoire d'Informatique Fondamentale de Marseille, France
Rozane Rebechi	Federal University Rio Grande do Sul, Brazil
M[a] Ángeles Recio Ariza	University of Salamanca, Spain
Irene Renau	The Pontifical Catholic University of Chile, Chile
Omid Rohanian	University of Wolverhampton, UK
Ute Römer	Georgia State University, USA
Leonor Ruiz Gurillo	University of Alicante, Spain
Agata Savary	François Rabelais University, France
Miriam Seghiri Domínguez	University of Malaga, Spain
Julia Sevilla Muñoz	Complutense University of Madrid, Spain
Kathrin Steyer	Institute of German Language, Germany
Joanna Szerszunowicz	University of Bialystok, Poland
Shiva Taslimipoor	University of Wolverhampton, UK
Yukio Tono	Tokyo University of Foreign Studies, Japan
Cornelia Tschichold	Swansea University, UK
Agnès Tutin	University of Stendhal, France
Aline Villavicencio	Federal University of Rio Grande do Sul, Brazil, and University of Essex, UK
Tom Wasow	Stanford University, USA
Eric Wehrli	University of Geneva, Switzerland
Juan Jesús Zaro Vera	University of Malaga, Spain
Michael Zock	French National Centre for Scientific Research, France

Additional Reviewers

Rocío Caro Quintana	University of Wolverhampton, UK
Souhila Djabri	University of Alicante, Spain
Anna Feherova	University of Wolverhampton, UK
Emma Franklin	Lancaster University, UK
Le An Ha	University of Wolverhampton, UK
Patrick Hanks	University of Wolverhampton, UK
Maria Kunilovskaya	University of Wolverhampton, UK
María Araceli Losey León	University of Cadiz, Spain

Carmen Merino Ferrada	University of Cadiz, Spain
Emad Mohamed	University of Wolverhampton, UK
Carmen Noya Gallardo	University of Cadiz, Spain
Encarnación Núñez	University of Wolverhampton, UK
Alistair Plum	University of Wolverhampton, UK
Maria Stasimioti	Ionian University, Greece
Victoria Yaneva	University of Wolverhampton, UK
María Ángeles Zarco Tejada	University of Cadiz, Spain

Keynote Speakers Main Conference

Sylviane Granger	Université Catholique de Louvain, Belgium
Miloš Jakubíček	Lexical Computing and Masaryk University, Czech Republic
Natalie Kübler	Paris Diderot University, France
Kathrin Steyer	Institute of German Language, Germany
Aline Villavicencio	Federal University of Rio Grande do Sul, Brazil, and University of Essex, UK

Invited Speakers of the MUMTTT 2019 Workshops

Jean-Pierre Colson	Université Catholique de Louvain, Belgium
Aline Villavicencio	Federal University of Rio Grande do Sul, Brazil, and University of Essex, UK

Organising Committee

Rosario Bautista Zambrana	University of Malaga, Spain
Rocío Caro Quintana	University of Wolverhampton, UK
Isabel Durán Muñoz	University of Malaga, Spain
Sandra Elfiky	University of Wolverhampton, UK
Javier Alejandro Fernández Sola	University of Malaga, Spain
Mahmoud Gaber	University of Malaga, Spain
Rut Gutiérrez Florido	University of Malaga, Spain
Carlos Manuel Hidalgo Ternero	University of Malaga, Spain
Suman Hira	University of Wolverhampton, UK
Pavlina Krasteva	University of Wolverhampton, UK
Maria Kunilovskaya	University of Wolverhampton, UK
Francisco Javier Lima Florido	University of Malaga, Spain
Gema Lobillo Mora	University of Malaga, Spain
Araceli Losey León	University of Malaga, Spain
Desiré Martos García	University of Malaga, Spain
Luis Carlos Marín Navarro	University of Malaga, Spain

Juan Pascual Martínez Fernández	University of Malaga, Spain
Sara Može	University of Wolverhampton, UK
Míriam Pérez Carrasco	University of Malaga, Spain
Alistair Plum	University of Wolverhampton, UK
Tharindu Ranasinghe Hettiarachchige	University of Wolverhampton, UK
Fernando Sánchez Rodas	University of Malaga, Spain
Anastasia Taramigou	University of Malaga, Spain
Shiva Taslimipoor	University of Wolverhampton, UK
Gergana Zyumbilska	University of Wolverhampton, UK

Association for Computational Linguistics (Bulgaria)

Nikolai Nikolov
Ivelina Nikolova

Sponsors

EUROPHRAS

EUROPHRAS
EUROPÄISCHE GESELLSCHAFT FÜR PHRASEOLOGIE

Sketch Engine

Sketch Engine

University of Wolverhampton

UNIVERSITY OF WOLVERHAMPTON

University of Malaga

Antequera

Málaga Convention Bureau

Contents

The Grammatical Environment of Intensifier-Noun Collocations: Insights from Lexical Priming Theory 1
Moisés Almela-Sánchez and Pascual Cantos-Gómez

Towards a Typology of Microsyntactic Constructions 15
Tania Avgustinova and Leonid Iomdin

'*Alla finfine sono daccordo*': A Corpus-Based Case Study on Italian Adverbial Phrases Grammaticalization 31
Lucia Busso and Margherita Castelli

On the Structural Disambiguation of Multi-word Terms 46
Melania Cabezas-García and Pilar León-Araúz

Phraseology in Specialised Language: A Contrastive Analysis of Mitigation in Academic Papers 61
María Luisa Carrió-Pastor

"Sword" Metaphors in Nikita Khrushchev's Political Rhetoric 73
Elena V. Carter

Multiword Units and N-Grams Naming FEAR in the Israel-Corpus 86
Carolina Flinz

Slovene Multi-word Units: Identification, Categorization, and Representation 99
Polona Gantar, Jaka Čibej, and Mija Bon

Weighted Compositional Vectors for Translating Collocations Using Monolingual Corpora 113
Marcos Garcia, Marcos García-Salido, and Margarita Alonso-Ramos

Phraseological Variation in Spanish Academic Writing 129
Marcos García Salido

Identifying Lexical Bundles for an Academic Writing Assistant in Spanish 144
Marcos García Salido, Marcos Garcia, and Margarita Alonso-Ramos

Do Online Resources Give Satisfactory Answers to Questions About Meaning and Phraseology? 159
Patrick Hanks and Emma Franklin

Translating Manipulated Idioms (EN>ES) in the Word Sketch Scenario 173
Carlos Manuel Hidalgo-Ternero

Nominal Collocations in Scientific English: A Frame-Semantic Approach . . . 187
Eva Lucía Jiménez-Navarro

Effects of Statistical Learning Ability on the Second Language Processing of Multiword Sequences 200
Elma Kerz and Daniel Wiechmann

Converse Phrasemes and Collocations in Czech: The Case of *dát* 'give' and *dostat* 'receive' 215
Marie Kopřivová

Automatic Identification of Academic Phrases for Czech 227
Dominika Kováříková and Oleg Kovářík

Translation Correspondences of Digressive Discourse Markers in English and Spanish: A Corpus-Based Study 239
Julia Lavid López

Context-Induced Reinterpretation of Phraseological Verbs: Phrasal Verbs in Late Modern English 253
Ljubica Leone

Adjectivation of Attributive Nouns in French and Spanish: A Corpus-Based Study of NOUN+clé/clave and NOUN+éclair/relámpago Expressions 268
François Maniez

MBLA Social Corpus: Multipurpose Multidimensional Corpus on Cyber-Language 283
Álvaro L. Maroto Conde and Manuel Bermúdez Vázquez

A Cognitive Modeling Approach on Ironical Phraseology in Twitter 299
Beatriz Martín Gascón

Profiling Idioms: A Sociolexical Approach to the Study of Phraseological Patterns 315
Sara Može and Emad Mohamed

Typical Phraseological Units in Poetic Texts: Frequent Concepts in Works of Poetry and How MWUs Lack Synonyms 330
Michael Pace-Sigge

Semantic Prosody in Middle Construction Predicates: Exploring Adverb + Verb Collocation in Middles 345
Macarena Palma Gutiérrez

Classification of the Combinatorial Behavior of Verbs in the Marketing Domain 360
Ivanka Rajh and Larisa Grčić Simeunović

"You Took the Word Out of My Mouth": A Morphosyntactic and Semantic Analysis of a Phraseological Lexicon of Colombian Spanish 375
José Luis Rojas Díaz and Juan Manuel Pérez Sánchez

Translating Emotional Phraseology: A Case Study 391
J. Agustín Torijano and Maria Ángeles Recio

The Difficult Identification of Multiworld Expressions: From Decision Criteria to Annotated Corpora 404
Agnès Tutin and Emmanuelle Esperança-Rodier

The Portrait of Dorian Gray: A Corpus-Based Analysis of Translated Verb + Noun (Object) Collocations in Peninsular and Colombian Spanish . . . 417
M.ª Victoria Valencia Giraldo and Gloria Corpas Pastor

Phraseological Sequences Ending in *of* in L2 Novice Academic Writing 431
Kateřina Vašků, Gabriela Brůhová, and Denisa Šebestová

Author Index 445