Gloria Corpas Pastor
Ruslan Mitkov (Eds.)

LNAI 13528

Computational and Corpus-Based Phraseology

4th International Conference, Europhras 2022
Malaga, Spain, 28–30 September, 2022
Proceedings

Springer

Lecture Notes in Artificial Intelligence 13528

Subseries of Lecture Notes in Computer Science

Series Editors

Randy Goebel
University of Alberta, Edmonton, Canada

Wolfgang Wahlster
DFKI, Berlin, Germany

Zhi-Hua Zhou
Nanjing University, Nanjing, China

Founding Editor

Jörg Siekmann
DFKI and Saarland University, Saarbrücken, Germany

More information about this subseries at https://link.springer.com/bookseries/1244

Gloria Corpas Pastor · Ruslan Mitkov (Eds.)

Computational and Corpus-Based Phraseology

4th International Conference, Europhras 2022 Malaga, Spain, 28–30 September, 2022 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 978-3-031-15924-4 ISBN 978-3-031-15925-1 (eBook)
https://doi.org/10.1007/978-3-031-15925-1

LNCS Sublibrary: SL7 – Artificial Intelligence

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2022
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

The corpus-based and computational study of phraseology is an important area relevant to the research and practical work of translators, interpreters, terminologists, lexicographers, language instructors, and learners, to mention a few. Computational phraseology has been gaining in prominence in recent years with the computational treatment of multiword expressions essential for a number of natural language processing and translation technology applications. The importance of this field has been evidenced in a number of influential publications including, but not limited to, Constant et al. (2017), Mitkov et al. (2018), Corpas and Colson (2020), and Ramish and Villavicencio (2022).

The Europhras conference series 'Computational and Corpus-based Phraseology' has emerged as one of the leading conferences covering this topic. We were delighted that after two years of restrictions due to COVID-19, we were again able to offer a face-to-face Europhras 2022. The volume you are holding features a selection of regular papers accepted after rigorous reviewing and presented at the latest edition of this popular conference, held during 28–30 September, 2022, in Malaga, Spain.

The conference was jointly organised by the University of Malaga (Research Group in Lexicography and Translation), Spain, the University of Wolverhampton (Research Group in Computational Linguistics), UK, and the Association for Computational Linguistics, Bulgaria, and was sponsored by Sketch Engine the University of Malaga, and Multilingual. Europhras 2022 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 interpreting, phraseology in language teaching and learning, phraseology in specialised languages, phraseology in lexicography, cognitive approaches to phraseology, phraseology and figurative language, phraseology in machine translation, and computational treatment of multiword expressions in general. 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 to present 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 or additional reviewers. The conference contributions were authored by a total of 75 scholars from 17 different countries: Brazil, Bulgaria, Colombia, Croatia, the Czech Republic, France, Indonesia, Iraq, Italy, Poland, Portugal, Russia, Slovakia, Spain, Sri Lanka, Sweden, and United Kingdom. These figures attest to the truly international nature of Europhras 2022.

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 2022 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 (Aline Villavicencio, Jean Pierre Colson, Miloš Jakubíček, and María del Carmen Mellado Blanco), the invited speakers of the accompanying workshop on Multiword Units in Machine Translation and Translation Technology (MUMTTT), and the tutorial speakers from Lexical Computing. Words of gratitude go to our sponsors and collaborators: Sketch Engine, Multilingual, the Research Institute of Multilingual Language Technologies of the University of Malaga, the Translation and Interpreting Department of the University of Malaga, and D.O.P Antequera - DCOOP.

Last but not least, we would like to use this paragraph to acknowledge the members of the Organising Committee and the Programme Committee of both the main conference and the MUMTT workshop, 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, Javier Alejandro Fernández Sola, Maria Kunilovskaya, and Desiré Martos García.

September 2022

Gloria Corpas Pastor
Ruslan Mitkov

References

Constant M., Eryiğit G., Monti J., Van Der Plas L., Ramisch C., Rosner, M. and A. Todirascu. 2017. "Multiword Expression processing: a survey". *Computational Linguistics*, 43(4). 837–892.

Corpas Pastor, G. and J.P. Colson (Eds.). 2020. *Computational Phraseology.* [IVITRA Research in Linguistics and Literature 24]. Amsterdam: John Benjamins.

Mitkov, R., Monti, J., Corpas Pastor, G. and V. Seretan (Eds). 2018. *Multiword Units in Machine Translation and Translation Technology.* [Current Issues in Linguistic Theory 341]. Amsterdam: John Benjamins

Ramish, C. and A. Villavicencio. 2022. "Computational treatment of multiword expressions". In Mitkov R. (Ed) *The Oxford Handbook of Computational Linguistics.* Oxford: Oxford University Press.

Organisation

Europhras 2022 was jointly organised by the University of Malaga (Research Group in Lexicography and Translation), Spain, the University of Wolverhampton (Research Group in Computational Linguistics), UK, and the Association for Computational Linguistics, Bulgaria.

Conference Chairs

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

Program Committee

Margarita María Alonso Ramos	University of A Coruña, Spain
M. Belén Alvarado Ortega	University of Alicante, Spain
Verginica Barbu Mititelu	Research Institute for Artificial Intelligence, Romanian Academy, Romania
Ignacio Bosque	Complutense University of Madrid, Spain
María Luisa Carrió-Pastor	Polytechnic University of Valencia, Spain
Anna Čermáková	University of Cambridge, UK
Parthena Charalampidou	Aristotle University of Thessaloniki, Greece
Ken Church	Baidu, USA
Jean-Pierre Colson	Université Catholique de Louvain, Belgium
Dmitrij Dobrovolskij	Russian Language Institute, Russia
Peter Ďurčo	University of St. Cyril and Methodius of Trnava, Slovakia
Natalia Filatkina	University of Hamburg, Germany
Elizaveta Goncharova	National Research University Higher School of Economics and Artificial Intelligence Research Institute (AIRI), Russia
María Isabel González Rey	University of Santiago de Compostela, Spain
Stefan Gries	University of California, Santa Barbara, USA
Enrique Gutiérrez Rubio	Palacký University Olomouc, Czech Republic
Kleanthes K. Grohmann	University of Cyprus, Cyprus
Amal Haddad Haddad	University of Granada, Spain
Miloš Jakubíček	Sketch Engine, Czech Republic
Eva Lucía Jiménez-Navarro	University of Cordoba, Spain
Cvetana Krstev	University of Belgrade, Serbia

Natalie Kübler	Université Paris Cité, France
Maria Kunilovskaya	University of Wolverhampton, UK
Ljubica Leone	Lancaster University, UK
Óscar Loureda Lamas	Heidelberg University, Germany
Elvira Manero Richard	University of Murcia, Spain
Ramón Martí Solano	Université de Limoges, France
María del 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	University of Naples "L'Orientale", Italy
Esteban Tomás Montoro del Arco	University of Granada, Spain
Inés Olza Moreno	University of Navarra, Spain
Adriane Orenha Ottaiano	São Paulo State University, Brazil
Antonio Pamies Bertrán	University of Granada, Spain
Rozane Rebechi	Federal University of Rio Grande do Sul, Brazil
Mª Ángeles Recio Ariza	University of Salamanca, Spain
Ute Römer	Georgia State University, USA
Leonor Ruiz Gurillo	University of Alicante, Spain
Kathrin Steyer	University of Mannheim, Germany
Joanna Szerszunowicz	University of Bialystok, Poland
Yukio Tono	Tokyo University of Foreign Studies, Japan
Agnès Tutin	Université Grenoble Alpes, France
Aline Villavicencio	Federal University of Rio Grande do Sul, Brazil, and University of Sheffield, UK
Tom Wasow	Stanford University, USA
Eric Wehrli	University of Geneva, Switzerland
Michael Zock	Laboratoire d'Informatique Fondamentale de Marseille, France

Additional Reviewers

Dayana Abuin Rios	University of Wolverhampton, UK
Rocío Caro Quintana	University of Wolverhampton, UK
Isabel Durán	University of Malaga, Spain
Richard Evans	University of Wolverhampton, UK
Emma Franklin	University of Wolverhampton, UK
Carlos Manuel Hidalgo-Ternero	University of Malaga, Spain
Nieves Jiménez Carra	University of Malaga, Spain
Alfiya Khabibullina	University of Wolverhampton, UK
Lilit Kharatyan	University of Wolverhampton, UK
Ruslan Mitkov	University of Wolverhampton, UK
Daria Sokova	University of Wolverhampton, UK

Keynote Speakers

Aline Villavicencio	Federal University of Rio Grande do Sul, Brazil, and University of Sheffield, UK
Jean-Pierre Colson	Université Catholique de Louvain, Belgium
María del Carmen Mellado Blanco	University of Santiago de Compostela, Spain
Miloš Jakubíček	Sketch Engine, Czech Republic

MUMTTT 2022 Workshop Chairs

Gloria Corpas Pastor	University of Malaga, Spain
Ruslan Mitkov	University of Wolverhampton, UK
Johanna Monti	University of Naples "L'Orientale", Italy
Maria Pia di Buono	University of Naples "L'Orientale", Italy

MUMTTT 2022 Program Committee

Giuseppe Attardi	University of Pisa, Italy
Verginica Barbu Mititelu	Institute for Artificial Intelligence, Romanian Academy Research, Romania
Jean-Pierre Colson	Université Catholique de Louvain, Belgium
Anna Beatriz Dimas Furtado	University of Wolverhampton, UK
Federico Gaspari	University for Foreigners "Dante Alighieri", Italy
Amal Haddad Haddad	University of Granada, Spain
Philipp Koehn	Johns Hopkins University, USA
Judyta Mężyk	Université Paris-Est Créteil, France, and University of Silesia in Katowice, Poland
Pavel Pecina	Charles University in Prague, Czech Republic
Éric Poirier	Université du Québec à Trois-Rivières, Canada
Carlos Ramisch	Aix-Marseille Université, France
Max Silberztein	Université de Franche-Comté, France
Kathrin Steyer	University of Mannheim, Germany
Beata Trawinski	University of Mannheim, Germany
Agnès Tutin	Université Grenoble Alpes, France

EUROPHRAS 2022 Organizing Committee

University of Malaga (Spain)

María de la Presentación Aguilera Crespillo
Marta Alcaide Martínez
Rosario Bautista Zambrana
Isabel Durán Muñoz

Javier Alejandro Fernández Sola
Mahmoud Gaber
Rut Gutiérrez Florido
Carlos Manuel Hidalgo-Ternero
Hanan Saleh Hussein
Adriana Iglesias Lara
Francisco Javier Lima Florido
Gema Lobillo Mora
Araceli Losey León
Jorge Lucas Pérez
Luis Carlos Marín Navarro
Desiré Martos García
Laura Noriega Santiáñez
Laura Parrilla Gómez
Míriam Pérez Carrasco
Encarnación Postigo Pinazo
María del Pilar Rodríguez Reina
Juan Antonio Sánchez Muñoz
Fernando Sánchez Rodas
Míriam Seghiri Domínguez
Cristina Toledo Báez

University of Wolverhampton (UK)

Dayana Abuin Rios
Isuri Anuradha
Anastasia Bezobrazova
Rocío Caro Quintana
Ana Isabel Cespedosa Vázquez
Amal El Farhmat
Suman Hira
Alfiya Khabibullina
Lilit Kharatian
Maria Kunilovskaya
Gabriela Llull
Kamshat Saduakassova
Kanishka Silva
Daria Sokova

Association for Computational Linguistics (Bulgaria)

Nikolai Nikolov

Europhras 2022 Secretary

Desiré Martos García	University of Malaga, Spain

Webmaster

Javier Alejandro Fernández Sola	University of Malaga, Spain

START Managers

Rocío Caro Quintana	University of Wolverhampton, UK
Maria Kunilovskaya	University of Wolverhampton, UK

Editorial Team

Gabriela Llull	University of Wolverhampton, UK
Dayana Abuin Rios	University of Wolverhampton, UK
Amal El Farhmat	University of Wolverhampton, UK
Daria Sokova	University of Wolverhampton, UK

Student Volunteers (University of Malaga, Spain)

Luis Carlos Marín Navarro (Coordinator)
Ángela Alguacil Suárez
Alba Bandera García
Ana Boza Pardo
Gabriella D'ottavio Simón
Camila Geijo Islas
Patricia Gómez Manzano
Alejandro Gutiérrez Usero
Marta Olivera Moreno
Sandra Raimundo Sánchez
Ana Ranchal Serrano
María de la Cabeza Rodríguez Martínez
Bárbara Ruiz Merino
Rocío Serrano Rey
María Utrera Muñoz

MUMTTT 2022 Organizing Committee

Gennaro Nolano	University of Naples "L'Orientale", Italy
Giulia Speranza	University of Naples "L'Orientale", Italy
Khadija Ait ElFqih	University of Naples "L'Orientale", Italy

Organizing Institutions

LEXYTRAD

Research Group in Computational Linguistics

Bulgarian Association for Computational Linguistics

Partners

University of Malaga

Sponsors

Sketch Engine

Multilingual

Research Institute of Multilingual Language Technologies

Department of Translation and Interpreting, University of Malaga

Antequera

Contents

Constructing the Digital Proverbial Thesaurus: Theoretical and Methodological Implications 1
Melita Aleksa Varga and Kristina Feldvari

Lexical Semantic Mind Maps Based on Collocations as a Tool for Teaching Vocabulary: A Case Study 16
María Auxiliadora Barrios Rodríguez

Teaching and Learning French Formulas 31
Elena Berthemet

Multi-word Term Translation: A Student-Centered Pilot Study 47
Sandra Bullón and Pilar León-Araúz

Interpreting Tomorrow? How to Build a Computer-Assisted Glossary of Phraseological Units in (Almost) No Time 62
Gloria Corpas Pastor

Readability and Communication in Machine Translation of Arabic Phraseologisms into Spanish 78
Mohamed El-Madkouri and Beatriz Soto Aranda

Detecting Bajan Phraseology: A Metalexicographic Analysis 90
Cristiano Furiassi

Image Schemas and Image Schematic Complexes: Enhancing Neural Machine Translation Networks 105
Amal Haddad Haddad

Reassessing *gApp*: Does MWE Discontinuity Always Pose a Challenge to Neural Machine Translation? 116
Carlos Manuel Hidalgo-Ternero and Xiaoqing Zhou-Lian

Make + Adjective Combinations During the Years 1850–1999: A Corpus-Based Investigation 133
Ljubica Leone

Metonymy in Spanish/L2 Teaching: A Cognitive Analysis of Color Idioms and Their Inclusion in the Córdoba Project Database 146
Beatriz Martín-Gascón

Bootstrapping a Lexicon of Multiword Adverbs for Brazilian Portuguese 160
Izabela Müller, Nuno Mamede, and Jorge Baptista

Figurative Expressions with Verbs of Ingesting in Croatian 175
Jelena Parizoska and Jelena Tušek

Some Insights on a Typology of French Interactional Prefabricated Formulas in Spoken Corpora 190
Marie-Sophie Pausé and Agnès Tutin

Specialized Idioms: From LGP to LSP Phraseological Paradigm 206
José Luis Rojas Díaz, Juan Manuel Pérez Sánchez, and Alejandro Arroyave Tobón

Author Gender Identification for Urdu Articles 221
Raheem Sarwar

Author Index 237