

Preface

Computational and Corpus-Based Phraseology: Recent Advances and Interdisciplinary Approaches

As the late and inspiring John Sinclair (1991, 2007) observed, knowledge of vocabulary and grammar is not sufficient for someone to express himself/herself idiomatically or naturally in a specific language. One has to have the knowledge and skill to produce effective and naturally phrased utterances, which are often based on phraseological units (the idiom principle). This is in contrast to the traditional assumption or open choice principle that lies at the heart of generative approaches to language. As Pawley and Syder (1983) stated more than three decades ago, the traditional approach cannot account for nativelike selection (idiomaticity) or fluency.

Language is indeed phraseological and phraseology is the discipline that studies phraseological units (PUs) or their related concepts referred to (and regarded largely synonymous) by scholars as multiword units, multiword expressions (MWEs), fixed expressions, set expressions, phraseological units, 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 that 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 (1977) observes that the number of MWEs in a speaker's lexicon is of the same order of magnitude as the number of single words (Jackendoff 1977). 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) state that they are overwhelmingly present in terminology and 41% of the entries in WordNet 1.7 are reported to be MWEs.

PUs do not play a crucial role only in the computational treatment of natural languages. 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, which are, for the time being, unfortunately still largely unexplored.

This volume features a selection of papers written by the invited speakers as well as regular papers presented at the international conference “Computational and Corpus-Based Phraseology: Recent Advances and Interdisciplinary Approaches” (Europhras 2017). The conference, which is organised jointly by the European Association of Phraseology (Europhras) and the Research Institute in Information and Language Processing of the University of Wolverhampton, and sponsored by Europhras, the Sketch Engine, ELRA and the University of Wolverhampton, provides the perfect opportunity for researchers to present their work, fostering interaction and collaboration between scholars working in disciplines as diverse as natural language processing, translation, terminology, lexicography, languages learning, teaching and assessment, and cognitive science, to name only a few. I organised the volume thematically into the following sections, which demonstrate the breath of the topics represented at Europhras 2017: (1) Keynote and Invited Papers, (2) Phraseology in Translation and Contrastive Studies, (3) Lexicography and Terminography, (4) Exploitation of Corpora in Phraseological Studies, (5) Development of Corpora for Phraseological Studies, (6) Phraseology and Language Learning, (7) Cognitive and Cultural Aspects of Phraseology, (8) Theoretical and Descriptive Approaches to phraseology, and (9) Computational Approaches to Phraseology. In fact, the variety of topics at Europhras 2017 is even more remarkable if we take into account other conference presentations that are not included in this volume – in addition to the regular papers, the conference also featured short papers and posters, which are published separately as e-proceedings with ISBN and DOI numbers assigned to every contribution.

Every submission to the conference was evaluated by three reviewers – i.e., members of the Programme Committee consisting of 46 scholars from 23 different countries, or 12 additional reviewers from eight countries, who were recommended by the Programme Committee. The conference contributions were authored by a total of 91 scholars from 24 different countries. These figures attest to the truly international dimension of Europhras 2017.

I would like to thank everyone who made this truly interdisciplinary and international event possible. I would like to start by thanking all colleagues who submitted papers to Europhras 2017 and travelled to London to attend the event. I am 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 – both the keynote speakers of the main conference (Ken Church, Gloria Corpas, Dmitrij Dobrovolskij, Patrick Hanks, Miloš Jakubiček) and the invited speakers of the two accompanying workshops (Carlos Ramish and Jean-Pierre Colson). Words of gratitude go to our sponsors – Europhras, the Sketch Engine, ELRA, and the University of Wolverhampton.

Last but not least, I would like to use this paragraph to acknowledge the members of the Organising Committee, who worked very hard during the last 12 months and whose dedication and efforts made the organisation of this event possible. I would like to mention (in alphabetical order) the following colleagues whom I would like to highlight for competently carrying out numerous organisational tasks and being ready

to step in and support the organisation of the conference whenever needed. My big thank you goes out to Amanda Bloore, Martina Cotella, Arianna Fabbri, April Harper, Sara Moze, Nikolai Nikolov, Ivelina Nikolova, Rocío Sánchez González, Andrea Silvestre Baquero, Shiva Taslimipoor, and Victoria Yaneva.

November 2017

Ruslan Mitkov

Computational and Corpus-Based Phraseology
Second International Conference, Europhras 2017,
London, UK, November 13-14, 2017, Proceedings
Mitkov, R. (Ed.)
2017, XV, 463 p., Softcover
ISBN: 978-3-319-69804-5