

Book reviews

Gloria Corpas Pastor, Jean-Pierre Colson (eds.): *Computational Phraseology* (IVITRA Research in Linguistics and Literature 24). Amsterdam & Philadelphia: John Benjamins, 2020. 327 pp. ISBN: 978-90-272-0535-3 / E-ISBN: 978-90-272-6139-7.

Within the IVITRA series (John Benjamins), the present publication bears witness to the recent but intensive efforts to consolidate the emerging discipline of computational phraseology. The wealth of approaches, research topics and languages covered in this volume does indeed illustrate the hectic activity in a field where, twenty years after Sag and Baldwin's prolific publication (e.g. Sag et al. 2002), researchers are still strenuously endeavouring to make multiword expressions (MWEs) become less of a *pain in the neck* for natural language processing systems. Syntactic anomaly, non-compositionality, ambiguity, discontinuity, variability, overlap, nesting... many are the challenges that still need to be addressed for the optimal processing of ubiquitous MWEs (Constant et al. 2017; Ramisch and Villavicencio 2018; Ramisch et al. 2018).

In its conception, Prof. Corpas Pastor and Prof. Colson are to be lauded for successfully bringing together key researchers both in natural language processing and phraseology, in order to demonstrate that there is much to gain from the cross-fertilisation of these convergent fields. In this regard, the overall structure of this volume takes the form of sixteen heterogeneous chapters, a comprehensive miscellanea in terms of typologies of phraseological units and electronic resources under study. This volume in fact includes nearly everything from the very left of the phraseological continuum (i.e. *collocations*) to the very right (viz. *idioms* and *proverbs*), and from general and specialised corpora to machine translation, with the main objective of optimising MWE-aware NLP systems.

Against such a background, an important part of the volume is devoted to the computational analysis of very specific types of phraseological units such as monocollapsible words (i.e. words whose use is restricted to a scarce number of phrasemes), constructional phrasemes (i.e. semi-fixed expressions with certain slots to be filled), and collostructions (a portmanteau word combining the notions of *collocation* and *constructions*), to which little computational attention had hitherto been paid. The not so paradoxical instability of fixed expressions also comes to the fore with several in-depth analyses of the most suitable corpus approaches for the syntactical and lexical variations which may affect both the core components of the phraseological unit and other optional elements occurring within the lexical patterns. Further challenges arise for NLP, and more

specifically for machine translation, when the phraseological anisomorphism between two or more language systems prompts translation asymmetries, by which there is not always a one-to-one (or, phraseologically speaking, a many-to-many) word translation. This key issue could certainly not be neglected in the present exhaustive volume.

Additional sections of the publication are dedicated to advancing the debate on novel measures and tools for the automatic and semi-automatic processing of phrasemes. In this regard, different techniques are proposed and further elaborated such as the new metric for the automatic extraction of phraseology: the Corpus Proximity Ratio or CPR (Colson 2016), the MERGE (Multi-word Expressions from the Recursive Grouping of Elements) algorithm based on the progressive extension of adjacent bigrams according to lexical association strengths (Wahl and Gries 2018) as well as practical tools combined with corpora for the automatic extraction of phrasemes by statistical scores: the mwetoolkit (Ramish 2015), all of which opens up innovative avenues of research in the field of MWE processing.

What is more important: the size or the quality of corpora? Which corpora (comparable, parallel, monolingual...) are more convenient for the analysis of phraseology? How can the complex links between phraseological and semantic associations be best approached from a computational perspective? The answer to these and other key questions is also to be found along the pages of this comprehensive volume which well represents the breadth and width of the growing body of literature in computational phraseology.

Carlos Manuel Hidalgo-Ternero

Correspondence address: cmhidalgo@uma.es

References

- Colson, Jean-Pierre. 2016. Set phrases around globalization: an experiment in corpus-based computational phraseology. In Francisco Alonso Almeida, Ivalla Ortega Barrera, Elena Quintana Toledo & Margarita Esther Sánchez Cuervo (eds.), *Input a word, analyze the world. Selected approaches to corpus linguistics*, 141–152. Newcastle: Cambridge Scholars Publishing.
- Constant, Mathieu, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner & Amalia Todirascu. 2017. Multiword expression processing: A survey. *Computational Linguistics* 43(4). 1–92.
- Ramisch, Carlos. 2015. *Multiword expressions acquisition: A generic and open framework* (Theory and Applications of Natural Language Processing 16). Cham: Springer.
- Ramisch, Carlos & Aline Villavicencio. 2018. Computational treatment of multiword expressions. In Ruslan Mitkov (ed.), *Oxford Handbook of Computational Linguistics* (2nd edn). N. p.: Oxford University Press. DOI: 10.1093/oxfordhob/9780199573691.013.56.

- Ramisch, Carlos, Silvio Ricardo Cordeiro, Agata Savary, Veronika Vincze, Verginica Barbu Mititelu, Archana Bhatia, Maja Buljan, Marie Candito, Polona Gantar, Voula Giouli, Tunga Güngör, Abdelati Hawwari, Uxoa Iñurrieta, Jolanta Kovalevskaitė, Simon Krek, Timm Lichte, Chaya Liebeskind, Johanna Monti, Carla Parra Escartín, Behrang QasemiZadeh, Renata Ramisch, Nathan Schneider, Ivelina Stoyanova, Ashwini Vaidya & Abigail Walsh. 2018. Edition 1.1 of the PARSEME shared task on automatic identification of verbal multiword expressions. In *Proceedings of the joint workshop on linguistic annotation, multiword expressions and constructions (LAW-MWE-CxG-2018)*, Santa Fe, New Mexico, USA, August 25–26, 2018, 222–240. <https://www.aclweb.org/anthology/W18-4925.pdf> (accessed 10 May 2020)
- Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In Alexander Gelbukh (ed.), *Computational linguistics and intelligent text processing. CICLing 2002. Lecture notes in computer science*, 1–15. Berlin & Heidelberg: Springer.
- Wahl, Alexander & Stefan Th. Gries (2018). Multi-word expressions: A novel computational approach to their bottom-up statistical extraction. In Pascual Cantos-Gómez & Moisés Almela-Sánchez (eds.), *Lexical collocation analysis: advances and applications*, 85–109. Berlin & New York: Springer.