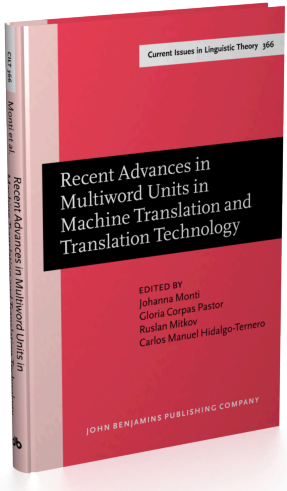
Current Issues in Linguistic Theory
366

Preface

Johanna Monti
Gloria Corpas Pastor
Ruslan Mitkov
Carlos Manuel Hidalgo-Ternero

https://doi.org/10.1075/cilt.366.preface



Pages vii–x of
Recent Advances in Multiword Units in Machine Translation and Translation Technology
Edited by Johanna Monti, Gloria Corpas Pastor, Ruslan Mitkov and Carlos Manuel Hidalgo-Ternero
[**Current Issues in Linguistic Theory**, 366] 2024. ix, 264 pp.

Preface

Johanna Monti, Gloria Corpas Pastor, Ruslan Mitkov
& Carlos Manuel Hidalgo-Ternero

The cross-fertilisation of phraseology with the fields of corpus linguistics, computational linguistics, and natural language processing, inter alia, has undoubtedly represented a decisive boost for the discipline, giving birth to computational phraseology. The relevance of this burgeoning field of study is clearly attested to by recent influential publications including but not limited to Constant et al. (2017), Mitkov et al. (2018), Corpas and Colson (2020), Corpas and Mitkov (2022) and Ramish and Villavicencio (2022). It is hence no wonder that the emergence of a discipline focusing on the appropriate computational treatment of multiword expressions (MWEs, also referred to as *multiword units* or *MWUs*) came like a breath of fresh air for language researchers and professionals such as translators, interpreters, terminologists, lexicographers, and language instructors and learners, to mention just a few.

In particular, the successful computational treatment of MWUs is essential for machine translation and translation technology; the inability to detect MWUs automatically may result in an incorrect (and even unfortunate) automatic translation and may jeopardise the performance of applications such as text summarisation, information retrieval, and sentiment analysis, among others. In this regard, their appropriate computational treatment is still far from being *a piece of cake*, since many of the challenges that these units pose for NLP systems are due to their quintessential features: syntactic anomaly, non-compositionality, diasystematic variation, ambiguity, and discontinuity. Therefore, there still remains a long road ahead until their optimal automatic processing.

To further pave this phraseological avenue, the study of multiword units in natural language processing has been gaining traction and in recent years the number of researchers and projects focusing on them has increased dramatically. Against this background, the present volume offers contributions from speakers at the 4th International Workshop on Multiword Units in Machine Translation and Translation Technology which was held as part of the EUROPHRAS 2019 conference in Malaga as well as speakers at the main conference itself. The conference and workshop contributions were authored by a total of 150 scholars from 27

different countries. These figures account for the truly international nature of the event.

The volume in front of you consists of 13 articles carefully selected after a rigorous reviewing process and they are grouped into two parts: 'Computational treatment of multiword units' and 'Corpus-based and linguistic studies in phraseology'. The contributions in this volume report not only the latest advances in computational and corpus-based phraseology but reiterate the importance of this discipline in machine translation and translation technology, and natural language processing in general as well as in corpus linguistics, lexicography, terminology, and linguistic studies more generally.

Regarding the first section, 'Computational treatment of multiword units', it is possible to observe, on the one hand, studies analysing, in qualitative and quantitative terms, the performance of some of the most robust neural machine translation systems on the market when it comes to accurately translating phraseology in the broad sense (Colson's inaugural chapter) and specialised phraseology (the chapters by Speranza & Monti, and by Kübler et al.). In this regard, different methods are proposed for improving NMT performance of both multiword units (more specifically, a text-preprocessing system of verb-noun idiomatic constructions, in the study by Hidalgo-Ternero & Corpas Pastor) and multiword terms (with the MWT bracketing protocol proposed by León-Araúz & Cabezas-García).

The second part of this volume, named 'Corpus-based and linguistic studies in phraseology', also includes a further study focusing on the bracketing of three-component MWTs (Rojas-Garcia's chapter) as well as additional methods of NMT-orientated MWU treatment, such as Palma Gutiérrez's colloconstructional analysis of MWUs for their disambiguation in machine translation, or di Buono et al. and Takahashi et al.'s phraseographic resources: *OntoLex-Lemon*, for Italian clitic verbal multiword expressions (MWEs), and *JMWEL*, for MWEs in Japanese, respectively.

The study of MWUs in specialized domains is thoroughly discussed in Jiménez-Navarro's chapter on verb collocations in the language of science and Brett et al.'s on negative-positive adjective pairing in EN/IT/PL travel journalism. Finally, regarding the analysis and categorisation of phraseology for the English language, the chapter by Martín-Gascón examines verbal irony in American-English tweets from a cognitive linguistics perspective, while Anna Frankhauser's study offers pioneering suggestions for a new model of functional phraseme categorization for applied purposes.

We would finally like to thank all authors who sent their contributions to this volume and the reviewers for finding time to review the submissions and provide

feedback. We would also like to seize this opportunity and thank John Benjamins for making the publication of this volume possible.

Johanna Monti
Gloria Corpas Pastor
Ruslan Mitkov
Carlos Manuel Hidalgo-Ternero

References

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

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

Corpas Pastor, G. & Mitkov, R. (Ed.). (2022). *Computational and corpus-based phraseology. Lecture notes in artificial intelligence*. Springer.

Mitkov, R., Monti, J., Corpas Pastor, G. & Seretan, V. (Eds). (2018). *Multiword units in machine translation and translation technology*. [Current Issues in Linguistic Theory 341]. John Benjamins.

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