Focus on computational and corpus-based phraseology at Europhras 2019

On 25-27 September 2019, a colleague from the Centre’s Terminology team attended the International Conference ‘Computational and Corpus-based Phraseology’ (Europhras 2019) in Málaga (Spain) and took note of the latest computerised approaches to phraseology.

The three-day programme featured presentations on a range of interdisciplinary topics. These included corpus-based, psycholinguistic and cognitive approaches to the study of phraseology, the computational treatment of multi-word expressions, as well as their practical applications in translation, lexicography and language learning, teaching and assessment.

Four international keynote speakers presented new research on computational phraseology.

Miloš Jakubíček, Chief Executive Officer (CEO) of Lexical Computing and a software developer of SketchEngine, gave a talk on automatic methods for extracting multi-word expressions (MWEs) based on standard properties such as fixedness, and the degree to which a multi-word item is frozen as a sequence of words.

Aline Villavicencio, Professor at the Department of Computer Science in the University of Sheffield (United Kingdom), presented an overview of the advances in the identification of MWEs, based on the various degrees of peculiarities they display, including lexical, syntactic, semantic and statistical information.

Natalie Kübler, Professor of Specialised Translation and Languages for Specific Purposes at the Paris Diderot University, showed how research and practice have evolved in specialised translation studies due to an increased focus on theoretical corpus-based translation.

Ruslan Mitkov, Professor at the University of Wolverhampton (United Kingdom), is a renowned expert in natural language processing and the sole Editor of The Oxford Handbook of Computational Linguistics (Oxford University Press) and Executive Editor of the Journal of Natural Language Engineering (Cambridge University Press). Professor Mitkov gave an interesting talk on the automatic translation of MWEs using comparable corpora.

EUROPHRAS 2019 also hosted the fourth Workshop on Multi-word Units in Machine Translation and Translation Technology (MUMTTT 2019), aimed at bringing together researchers working on Natural Language Processing (NLP) approaches based on the computational treatment of multi-word units.

At the EUROPHRAS 2019 conference, it became clear to all that phraseological data is available like never before thanks to new corpus-based technology and to new algorithms based on computational techniques to identify MWEs clusters. Corpora will continue to grow in the next years and phraseology studies must benefit from this approach. One of the main challenges for Machine Translation continues to be the recognition of phraseology so as to avoid literal translation from the analysis of compositional expressions in the source text. Therefore it is crucial for linguists and engineers to concentrate their efforts on the study and identification of MWEs to enable their automatic identification in large corpora.