Multi-word expression processing: a survey and a close-up of deverbal compounds

Multiword expressions (MWEs) are a class of linguistic forms spanning conventional word boundaries that are both idiosyncratic and pervasive. The structure of linguistic processing that depends on the clear distinction between words and phrases has to be re-thought to accommodate MWEs. In the first part of my talk, I will give a focused review of MWE processing and clarify the nature of interactions between MWE processing and downstream applications that give rise to different orchestration scenarios (how processing is timed with respect to underlying use cases). In the second part of my talk, I will move away from NLP processing and zoom into one type of MWE, namely deverbal compounds. I will show how simple machine learning techniques help to verify and inspire linguistic theory. Conversely, linguistic theory provides great inspiration for the use of alternative features for the processing of deverbal compounds in NLP.