Modelling the internal variability of MWEs

Malvina Nissim
Department of Linguistics and Oriental Studies
University of Bologna
Via Zamboni 33, 40126 Bologna, Italy
`malvina.nissim@unibo.it`

Abstract

The issue of flexibility of multiword expressions (MWEs) is crucial towards their identification and extraction in running text, as well as their better understanding from a linguistic perspective. If we project a large MWE lexicon onto a corpus, projecting fixed forms suffers from low recall, while an unconstrained flexible search for lemmas yields a loss in precision. In this talk, I will describe a method aimed at maximising precision in the identification of MWEs in flexible mode, building on the idea that internal variability can be modelled via so-called variation patterns. I will discuss the advantages and limitations of using variation patterns, compare their performance to that of association measures, and explore their usability in MWE extraction, too.

About the Speaker

Malvina Nissim is a tenured researcher in computational linguistics at the University of Bologna. Her research focuses on the computational handling of several lexical semantics and discourse phenomena, such as the choice of referring expressions, semantic relations within compounds and in argument structure, multiword expressions, and, more recently, on the annotation and automatic detection of modality. She is also a co-founder and promoter of the Senso Comune project, devoted to the creation of a common knowledge base for Italian via crowdsourcing. She graduated in Linguistics from the University of Pisa, and obtained her PhD in Linguistics from the University of Pavia. Before joining the University of Bologna she was a post-doc at the University of Edinburgh and at the Institute for Cognitive Science and Technology in Rome.