Explorations into translation-based paraphrasing and sense induction

Synonym and paraphrase resources extracted from parallel corpora are widely used in translation applications. They permit to abstract away from surface correspondences and obtain a generalisation across meanings that reduces sparseness and improves evaluation. In this talk, I will explore the semantics of translation-based paraphrases commonly used in multilingual tasks. I will demonstrate the need for further analysis to account for sense distinctions present in these resources and avoid erroneous judgments of semantic relatedness. I will then investigate the adequacy of cross-lingual sense representations in translations and present some avenues for a more effective modelling of semantics in multilingual applications.