Using translations into many languages to predict the compositionality of multiword expressions

Multiword expressions (MWEs) are combinations of words that exhibit some degree of idiosyncrasy, including verb-particle constructions such as "strike up", and noun compounds such as "gravy train". MWEs occupy a continuum of compositionality, whereby some expressions are non-compositional and have meanings that are largely unrelated to that of their component words (e.g., "ivory tower"), while others are compositional and their meanings can be easily inferred from that of their components (e.g., "climate change"), and still others lie somewhere in between (e.g., "silver screen"). In this talk I will present a method for predicting MWE compositionality that is applicable to many kinds of MWEs, and is not restricted to any particular language. The method is based on translations of the MWE and its component words into many languages, and then measuring the similarity between the translation of the MWE and the translations of each of its components in terms of string similarity and distributional similarity. I will present results on German and English noun compounds, and English verb-particle constructions, that demonstrate that exploiting information from translations into multiple languages improves over a monolingual model. On the two English datasets this method is an improvement on the previous state-of-the-art.