Word Association Measures for Finding Verb Sense Patterns.

Patrick Hanks, Ismail El Maarouf, Michael Oakes.

We use statistical measures proposed by Van de Cruys called specific correlation (SC) and specific interaction information (SII) to find multi-word expressions (MWE) in the BNC50 corpus. We show results for two kinds of MWE: a) subject-verb-object triples, and b) all triples consisting of a verb and two arguments. We compare our automatically extracted MWU lists against a "gold standard" list of idioms and phrasal verbs produced by the Pattern Dictionary of English Verbs (PDEV) project. In general, SC slightly outperformed SII and the baseline raw frequency.