

Multi-Word Units as a Model of Grammar

Oliver Mason¹

Abstract

Based on previous work on the automatic extraction of multi word units from text corpora I will be investigating how these units can be used to describe the structure of sentences. The result is similar to the pattern grammar approach by Hunston and Francis, though instead of patterns we are using multi word units as the fundamental unit of description. These units are of course not necessarily grammatically motivated or even 'well-formed', but illustrate the role that creativity and routine usage play in sentence formation.

We propose a model of sentence/utterance structure which moves away from a purely hierarchical approach to a multi-layered hybrid one that is mostly linear: routine elements are realised using multi word units (as prefabricated fragments), whereas creativity (which cannot easily be described through recurring elements) requires more 'open' ways of description, which are less lexicalised (such as pattern grammar, or even phrase structure grammar).

Based on an empirical analysis of a variety of text types we try to arrive at a syntactic model that is suitable for describing authentic sentences, making use of several layers with different formalisms governing each layer. This hybrid model aims to be more appropriate than the single formalisms in isolation.

¹ University of Birmingham
e-mail: O.Mason@bham.ac.uk