

**Doing Research in Applied Linguistics 2 / Independent Learning Association Conference**

**King Mongkut`s University of Technology, Thonburi, Thailand 2014**

**12-14 June 2014**

**John Blake**

**Corpus linguistics: Pitfalls and problems**

The advent of concordancing software has enabled linguists to search corpora and identify patterns that are unlikely to be discovered through introspective intuition. Corpus linguistics can be harnessed to investigate various aspects of language. These can include, for example, rhetorical structures, (e.g. moves and steps), phraseologies, such as frames (e.g. triples with fixed first and third words), and hidden meanings (e.g. *have a point* used to express partial agreement). However, a number of factors relating to the research methodology, such as representativity, sample size and balance of a corpus may reduce the validity of the results. Using a case study of an investigation into the structure and phraseology of scientific research abstracts, this paper describes the potential problems and pitfalls in the design, construction, annotation and analysis of a corpus. Solutions and workarounds that increase the reliability, validity and generalizability of results stemming from corpus-based or corpus-led studies will be suggested.