

Analysing a Corpus-based Semantic Investigation of English Dialects

Dr Debra Ziegeler & Ms Sarah Lee

School of English & Linguistics
University of Manchester
debra.ziegeler@man.ac.uk
sarah.lee@man.ac.uk
Fax: +44 (0)161-275 3256

The “Causatives in Singapore English Project” investigates cross-dialectal preferences for encoding caused agency (agents with [+ volition], [+ intention]) (Talmy 1979, 2000) causative events. In more studied dialects such as British English, a common strategy used to encode caused agency causatives are the *get* and *have* possessive periphrastic constructions, i.e., [NP1-GET/HAVE-NP2-STEM/INF/-ING]. We conducted a contrastive corpus-based analysis of various English dialects (British English, Singapore English, New Zealand English, and Indian English) in the first instance, to obtain a general quantitative picture of the frequency distribution of possessive periphrastics and of other strategies enlisted by the respective dialects. It was found that the periphrastics were present in all dialects but there were significant frequency and in some cases, distribution differences across dialects. We also found competing strategies, again in varying frequencies across the dialects, such as the resultative [NP1-GET/HAVE-NP2-PP], ‘conventionalised scenario’ constructions (Goldberg 1995), and other verbs such as *ask* and *tell* used in an implicative sense. We concluded that a corpus-based analysis as an initial tool of investigation was indeed valuable as a springboard for further empirical investigation.

Nevertheless, we also found some evidence to suggest that the design of a corpus itself can skew the statistics obtained. Our topic was concerned with semantic (and grammatical function) information of a particular type, and we found that caused agency, a salient property, was not equally distributed amongst the components found in each corpus. This was because the components were stratified across varied genres. Some contained written as well as spoken data, and others just the former. Some genres represented formal styles, others informal. Thus statistical testing strategies we used had to include deriving average comparative probabilities of particular components across dialects and determining the probability of occurrence of a particular causative strategy in a given semantic environment.