

Counting matters: Normalization and accountability

Alexandra D'Arcy

University of Victoria

From a methodological perspective, corpus linguistics and variationist sociolinguistics overlap in their focus on frequencies of use and probabilities in distribution. However, the quantitative model is fundamentally different in each field. Where corpus linguists rely on normalization, variationists working in the Labovian paradigm apply the principle of accountability (Labov 1972, 1982) and consider the feature of interest as well as all contexts where it could have occurred but did not. Such divergent approaches to what to count and how to count have analytical repercussions, affecting our view of the variable feature and how it works in language. This paper presents a methodological case study, comparing the results of corpus linguistic methods to variationist methods in the analysis of discourse *like*.

Discourse *like* is a ubiquitous feature of vernacular speech. It has drawn extensive scholarly attention (e.g. Ross & Cooper 1979; Underhill 1988; Romaine & Lange 1991; Andersen 2001), but notably little of this work has been carried out under the umbrella of variationist sociolinguistics (e.g. Dailey-O'Cain 2000). The greatest challenge is the infeasibility of 'delineating a universe of discourse for pragmatic devices which includes opportunities for potential occurrence as well as actual occurrences' (Stubbe & Holmes 1995:71). Discourse *like*, however, is used in a number of syntactically delimited contexts (Underhill 1988; Andersen 2001). This enables a unique approach to the analysis of *like*, one in which the variable context is circumscribed on structural grounds. Quantifying *like* in this way reveals that it is highly constrained by the syntax, occurs in specific positions among speakers of all ages, and has developed gradually and systematically, arriving at its current state through regular processes of language change (D'Arcy 2007, 2008).

The focus of the current analysis is methodological. The most thorough investigation of *like* in the corpus linguistic tradition is that of Andersen (2001), who makes a number of observations concerning the distribution of *like* (e.g. formulaic status of frequent collocations; high text frequency before noun phrases, verb phrases, and clauses; tendency to occur before lexical rather than functional material, i.e. the principle of lexical attraction; etc.). The quantificational results underpinning these claims are derived via normalized text frequency. Each of these phenomena is examined here within the variationist framework, and Andersen's normalized results are systematically assessed vis-à-vis those derived from adherence to the principle of accountability.

The results of this comparative analysis corroborate some of the findings presented in Andersen 2001. For example, *like* is more frequent in certain syntactic positions than in others, but the difference is more subtle than suggested by text frequency. Other corpus linguistic findings are not substantiated by variationist methods. The principle of lexical attraction finds little support, as do claims regarding certain collocational patterns. Thus, while it is argued that the quantificational model of variationist sociolinguistics allows greater detail in uncovering the grammatical patterns of use that are revealed by accountability, a method that combines corpus linguistic and variationist traditions permits observations that are not possible using just one approach.

References

- Andersen, G. 2001. *Pragmatic markers and sociolinguistic variation*. Amsterdam/Philadelphia: Benjamins.
- D'Arcy, A. 2007. *Like* and language ideology: Disentangling fact from fiction. *American Speech* 82: 386–419.
- D'Arcy, A. 2008. Canadian English as a window to the rise of *like* in discourse. In M. Meyer (ed.), *Focus on Canadian English*. Special issue of *Anglistik: International Journal of English Studies* 19: 125–140.
- Dailey-O'Cain, J. 2000. The sociolinguistic distribution of and attitudes toward focuser *like* and quotative *like*. *Journal of Sociolinguistics* 4: 60–80.
- Labov, W. 1972. *Sociolinguistic patterns*. Philadelphia: University of Philadelphia Press.
- Labov, W. 1982. Building on empirical foundations. In W.P. Lehmann and Y. Malkiel (eds.), *Perspectives on historical linguistics*. Amsterdam/Philadelphia: Benjamins. 17–92.
- Romaine, S. and D. Lange. 1991. The use of *like* as a marker of reported speech and thought: A case of grammaticalization in progress. *American Speech* 66: 227–279.
- Ross, J.R. and W.E. Cooper. 1979. *Like*, syntax. In W.E. Cooper & E.C.T. Walker (eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale: Erlbaum. 343–418.
- Stubbe, M. and J. Holmes. 1995. *You know, eh* and other 'exasperating expressions': An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language and Communication* 15: 63–88.
- Underhill, R. 1988. *Like is like*, Focus. *American Speech* 63: 234–246.