

**Sociopragmatic annotation:  
New directions and possibilities in Historical Corpus Linguistics**

Dawn Archer and Jonathan Culpeper  
Department of Linguistics and Modern English Language, Lancaster University  
{d.archer, j.culpeper}@lancaster.ac.uk

The bias of computer searches towards form (e.g. a letter or string of letters) can be a major difficulty for linguistic analyses of texts. In particular, how form relates to context in interactive texts tends to be overlooked. Corpus linguists have been seeking to improve our understanding of the relationships between forms (e.g. collocations, lexical networks, keywords); in other words, they focus on text and co-text. However, little work has considered context in terms of situational, sociological and cultural phenomena. Consequently, our aim in this paper is to demonstrate how a sophisticated annotation scheme can help bridge the gap between text and contexts.

Our work is located in the area of historical corpus linguistics, an area where there are particular difficulties relating to the retrieval of contextual factors (e.g. the possibility of asking speakers is not available!) and where little work has been done on annotation of any kind. Specifically, we will introduce an annotated sub-section of the Corpus of English Dialogues (CED), constructed by Jonathan Culpeper and Merja Kytö (Uppsala University). The Sociopragmatic Corpus, as it is called, covers a 120 year time span (1640-1760), and consists of more than 240,000 words of “authentic” and constructed speech drawn from two text-types: trial proceedings and drama. We will then describe how the annotation system that we have developed:

- (i) Accommodates the investigation of language set in various context(s) (for example, speaker/hearer relationships, social roles, and sociological characteristics such as gender), and
- (ii) Treats contexts as dynamic (cf. other annotation systems, such as the spoken sub-section of the BNC, which concentrate upon the relatively static characteristics of speakers).

More generally, this paper shows how computational linguistics can be used in pragmatics research.