The corpus in Corpus Pragmatics: decisions, data, and desiderata
Rachele De Felice and Svenja Adolphs

Pragmatics research increasingly makes use of corpus data and corpus linguistics analysis, rather than introspection or artificially elicited data (cf. e.g. Romero-Trillo 2008, Jucker et al. 2009, O'Keeffe et al. 2011). The process of acquiring and analysing data, however, is not always straightforward, especially since there are not many corpus resources developed specifically for pragmatic research.
 This talk addresses some methodological and theoretical issues encountered in corpus pragmatics research, taking as a case study a research project on speech acts in Business English. Issues include:
 How do we identify the corpora best suited to our needs? Do we need to create new ones or can we work with existing resources?
 How much data do we need? Does increasing the amount of data affect our conclusions?
 What kind of annotation is required?
 What tools best serve our analysis? Can natural language processing techniques enhance and support our investigations?
This presentation discusses how these issues have been addressed in the course of our research, and shows how this project has the potential for wide-ranging impact. As well as validating and expanding current research on pragmatics, it constitutes an informative study on how to bring together natural language processing and applied linguistics techniques (cf. e.g. De Felice and Deane 2009). Furthermore, its findings contribute to developing new teaching materials for English as a foreign language, supported by empirical evidence rather than artificially constructed scenarios.

 R. De Felice and P. Deane, 2009. Identifying speech acts in emails: Business English and non-native speakers. Paper presented at the Corpus Linguistics Conference, Liverpool
 A. Jucker, D. Schreier, and M. Hundt (eds), 2009. Corpora: Pragmatics and Discourse. Rodopi
 A. O'Keeffe, B. Clancy, and S. Adolphs, 2011. Introducing Pragmatics in Use. Routledge
 J. Romero-Trillo (ed.), 2008. Pragmatics and Corpus Linguistics: a Mutualistic Entente. De Gruyter