

John M. Kirk, Jeffrey L. Kallen, Orla Lowy and Anne Rooney
Queen's University, Belfast

Pragmatics, Prosody, and Syntax in Spoken Language Corpora

Orthographic transcription offers tremendous advantages for corpus linguistics, in allowing for efficient computerised searches for lexical, syntactic, and other features. Yet the traditional orthographic transcription fails to make any transparent accounting for the pragmatic aspects of spoken language, both because of the loss of prosodic information which is vital for the encoding and decoding of speakers' intent, and because there is no one-to-one match between syntax or lexicon and the pragmatic intentions of speakers. Following from the creation of the International Corpus of English in Ireland (ICE-Ireland), which has generated a 1-million word corpus using the purely orthographic transcription system shared with other ICE corpora, a further research project has taken on the task of enriching transcription with prosodic and pragmatic information. By adapting the Tones and Break Indices (TOBI) system (Silverman et al. 1992) for use within the corpus, and developing an annotation system from speech act theory and the analysis of discourse markers (Searle 1969, 1975; Schiffrin 1987, Stenström 1994, Aijmer 2002), we are in the process of developing an enriched ICE-Ireland corpus, which includes not only orthographic transcription but information on prosody and the pragmatic status of individual utterances and discourse markers. The poster will display our complete prosodic and pragmatic annotation scheme and discuss the many issues which we had to face as well as the solutions which we have adopted. By developing such a system, it becomes possible not only to describe spoken language in considerable detail and to show the relationship between prosody and the speech act, but to make comparisons which reveal socially significant variation in the marking of pragmatic intent.