

**Crowd-sourcing prosodic annotation:
How Big (& Noisy) Data can inform the linguistic analysis of prosody in natural discourse**

Jennifer Cole, University of Illinois

Investigations of prosody in the production or perception of speech depend on the availability of prosodic annotation that marks the prosodic features of an utterance—the prominences and boundaries. Major challenges for prosodic annotation arise due to: the ambiguity of acoustic cues to prosodic features, individual speaker differences in the phonological encoding and phonetic implementation of prosody, the simultaneous influence of top-down and signal-based cues to prosody, and a lack of attention to the criterion of meaningful contrast. These factors negatively impact the speed and reliability of prosodic annotation, even when performed by highly trained expert transcribers, and are especially problematic for annotation of spontaneous speech produced in meaningful communication, where prosody is notably rich and varied.

In this talk I discuss Rapid Prosody Transcription (RPT), a crowd-sourcing approach to the study of prosody that replaces the expert transcriber with ordinary, untrained individuals, and which embraces inter-transcriber variability in prosodic annotation as a source of information. I report findings from two studies of prosody in the Buckeye corpus of American English conversational speech, where RPT annotations are shown to be a good approximation of expert annotations using the Tones and Break Indices (ToBI) system. RPT annotations are also compared across different task conditions, where transcribers are instructed to focus on acoustic cues or meaning-based criteria for transcribing speech presented in auditory format, and in the annotation of “imagined” prosody with text-only presentation of transcribed conversation excerpts. Our findings show convergence among transcribers and across transcription methods in prosodic boundary annotation, but persistent variability in the identification of prominence. Variability in prominence annotation, which is highest for words with prenuclear prominence, is related to variability in the acoustic cues to prominence and to the divergence between prominence as judged from acoustic, meaning-based, or lexico-syntactic cues. Implications from these studies are discussed as they relate to the practice of prosodic annotation and theories of the linguistic encoding of prosody.