I like spontaneous speech data and I like standardised elicitation task data, but which is better? There's only one way to find out…

Will Barras
Linguistics and English Language, The University of Edinburgh
will@ling.ed.ac.uk

I am researching the relationship between rhoticity and *r*-sandhi in an area of Lancashire where there is both inter-speaker and intra-speaker variation in rhoticity. Recording spontaneous speech intuitively seems to be the best way of accessing the vernacular, which according to Labov "gives us the most systematic data for our analysis of linguistic structure" (1972: 208). However, while potential instances of coda-/r/ occur frequently in spontaneous speech, potential instances of intrusive-*r* occur far less frequently. Following approaches used by other researchers investigating these two linked phenomena (for example Hay and Maclagan forthcoming), I have designed elicitation tasks to generate more frequent tokens of intrusive-*r* than are obtained in "natural" spontaneous speech. In practical terms, I would need to record many hours of conversation to obtain a sufficient number of tokens of intrusive-*r* for convincing analysis to be possible. Furthermore, in order to investigate a full range of possible preceding vowels for *r*-sandhi, even more hours of spontaneous speech would be necessary. The use of a standardised elicitation task ensures that all speakers across the study produce a comparable set of tokens, which is important given that I am attempting a sociophonological analysis of the variation in my speakers' production of *r*-sandhi. This reflects comments by Hay and Maclagan about the nature of their elicited data: "we sacrificed a great deal of 'naturalness' in order to record the full paradigms. Participants were warned that some of the sentences they would be asked to read would be a bit weird, and they were asked just to humour us!" (2006: 5).

However, Bigham (2008) argues that elicited data are not merely a less-than-wholly-satisfactory workaround when dealing with relatively low-frequency phenomena: in fact they can reveal patterns that are masked in spontaneous speech data. In a study of a vowel merger in Illinois speech, he argues that "'more attentive speech' more accurately reflects phonological categories". Elicited data are, therefore, vital for considering dialectal phonological variation, and should not be regarded as inferior to spontaneous speech data.

In this paper, I will compare the spontaneous and elicited data for one of my participants in order to consider: (1) the difference in the numbers of tokens gathered in each type of data; (2) whether there are systematic differences in the realisations of tokens recorded; (3) why a combination of spontaneous speech and elicited speech is appropriate for my research.

References

Bigham, Douglas S., 2008. *On the Importance of Standardized Word List Data in Dialectology Research*. Methods in Dialectology XIII – University of Leeds.

Hay, Jennifer and Margaret Maclagan (forthcoming). Social and phonetic conditioners on the frequency and degree of intrusive /r/ in New Zealand English. To appear in Dennis Preston and Nancy Niedzielski (eds) *Methods in Sociophonetics*.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.