Scaling up reliably: The pros and cons of rigorous protocols

Suzanne Evans Wagner
Michigan State University

Quantitative discourse-pragmatic research confronts a multitude of methodological challenges (Pichler 2010). Given the highly situated nature of discourse-pragmatic variables, it can be difficult for analysts to generalize confidently across data from multiple contexts and sources, particularly if the object of inquiry is the variable's pragmatic *functions* rather than its *forms*. Yet comparisons are essential if we are to take a variationist sociolinguistic perspective on pragmatic function. For a given discourse-pragmatic variable, how are its functions distributed across linguistic environments, conversational contexts, demographic groups, dialects and generations? This knowledge can answer crucial questions about the transmission, incrementation and diffusion of the variable and about discourse-pragmatic change in general.

Practically speaking, then, we need coding protocols that are as reliable and reproducible as possible. They should consistently identify the same discourse-pragmatic functions of a variable, regardless of the data source or the stylistic/situational context. But this is a huge challenge on two fronts. First, style is an immensely problematized area of investigation in sociolinguistics that does not submit easily to one-size-fits-all coding schemes (Eckert & Rickford 2001). Second, identifying, e.g. a given discourse function (or multiple functions) of a token of *well* is not as straightforward as noting the preceding phonological environment of a stressed vowel.

Nonetheless, it is incumbent on us to at least try some one-size-fits-all coding methods for pragmatic function, even if this means narrowing the research scope in the initial testing phase. I will describe some efforts, using multiple corpora, to reliably identify and code English general extenders (*...and stuff, or anything, things like that* etc.) for their pragmatic functions (Wagner et al 2015). I will also report on more recent work that takes into account the stylistic contexts in which general extenders and their functions are embedded (Wagner et al fc, Gregersen & Barner-Rasmussen 2011). What is gained and what is lost when we prioritize the needs of large-scale quantitative analysis over smaller scale, more qualitative approaches? I'll evaluate the degree to which applying rigorous function and style coding protocols do and do not resolve these tensions.

References

Eckert, Penelope & John R. Rickford. 2001. *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Gregersen, Frans & Michael Barner-Rasmussen. 2011. The logic of comparability : On genres and phonetic variation in a project on language change in real time. *Corpus Linguistics and Linguistic Theory* 7(1). 7–36.

Pichler, Heike. 2010. Methods in discourse variation analysis: Reflections on the way forward. *Journal of Sociolinguistics* 14(5): 581–608.

Wagner, Suzanne Evans, Ashley Hesson & Heidi Little. forthcoming in 2016. Comparing referential general extender use across registers in American English speech. In Heike Pichler (ed.)

Discourse-Pragmatic Variation and Change in English: New Methods and Insights. Cambridge: Cambridge University Press.

Wagner, Suzanne Evans, Ashley Hesson, Kali Bybel & Heidi Little. 2015. Quantifying the referential function of general extenders in North American English. *Language in Society* 44(5).