

Vignette 2c Studying Difficult to Study Variables

J. Daniel Hasty

Though the gold standard for sociolinguistic research since the pioneering work of Labov (1966) has been the sociolinguistic interview, some morphosyntactic features and discourse/pragmatic features simply do not occur often enough (or even at all) in the traditional sociolinguistic interview. So, when attempting to analyze these hard to find linguistic variables where they naturally occur (rather than in artificial solicitation or through acceptability judgments), it helps to locate the discourse situation most likely to produce them. One such environment is the medical consultation, in which physicians interact with their patients in face-to-face settings. In this vignette I will use one hard to find morphosyntactic feature—the double modal, as in (1)—to illustrate some of the possibilities and drawbacks of using medical consultation to study hard to find variables.

- (1) a. Another thing we **might could** add on is the Neurontin.
 b. So you're kind of thinking about that, that might be something that **might would** help you?

Many methodological issues can arise when working with hard to find sociolinguistic variables that are not phonological, especially certain morphosyntactic or discourse/pragmatic variables. Some of these difficulties include establishing semantically or functionally equivalent variants, dealing with pragmatic constraints on the variable context, and finding enough tokens for statistical analysis (cf. Lavandera, 1978; Pichler, 2010; Wolfram, Vignette 2b). While some of these issues may ultimately lay with theoretical concerns as to the nature of the linguistic variable itself and its extension to areas other than the phonological (see Hasty, 2014), at least one way to partially overcome them is by looking in new places and with new expectations. If we open up our studies to non-traditional or novel corpora (see, for instance, Coats, Vignette 14b, on compiling a corpus of Twitter data), some of these hard to study or hard to find variables may present themselves.

One excellent corpus, containing over 45,000 fully transcribed and searchable audio recordings of doctor-patient consultations from across the U.S., is collected and maintained by Verilogue Inc. (see Kozloff & Barnett, 2006). While HIPAA compliance necessitates removal of certain identifiable information, the transcripts do provide a good deal of demographic detail, including the patient's age, gender, race, employment status, health insurance type, and medical information, as well as the doctor's gender, years in practice, medical specialty, and the location (state only) of the practice.

In 2012, my colleagues and I investigated the appearance of double modals in the Verilogue corpus (Hasty, Hesson, Wagner, & Lannon, 2012). We found 95 double modal tokens in the corpus, further illustrating that even when a favorable pragmatic situation is located, some variables are still quite low occurring. Additionally, while traditional

applications of the sociolinguistic variable would spend quite some time determining situations where the feature could have occurred but crucially did not—i.e., the so-called envelope of variation (see Wolfram, Vignette 2b)—some variables, especially some morphosyntactic variables and variables with complex pragmatic constraints like the double modal, pose the extremely difficult task of identifying a true situation of non-occurrence. Take, for example, stressed *BIN* in African American English (AAE), completive *done* in AAE and Southern English, and emphatic pronoun tags in Northern British varieties, as in (2).

- (2) a. She *BIN* running. (53a in Green, 2002, p. 55)
 ~*She has been running for a long time.*
 b. I *done* told you once.
 ~*I have already completely told you.*
 c. I don't like it *me*. (4j in Cheshire, Kerswill, & Williams, 2005, p. 159)
I don't like it myself (emphatic).

Variables like those in (2) present an issue for studies using traditional quantitative variationist methods because they lack strict semantic equivalence with another syntactic form or even a clearly identifiable alternative at any level of the grammar. This is exactly also the case for the double modal, which seems to have no clear other form or construction with which it alternates. That is, it cannot be said that (3a) or (3b) are alternate forms of the double modal in (3c): neither form provides the meaning encoded in (3c), which can best be described as limiting the possible worlds in which the speaker believes that the addressee should go to the store.

- (3) a. You might go to the store.
 b. You should go to the store.
 c. You *might should* go to the store.

Thus, at this point in our knowledge of these kinds of variables we are limited in what we can gather from quantitative studies, and we can never get true estimates of percentage of use of these features. There are things we can learn from the data if we view these kinds of variables slightly differently, however. Rather than attempting to determine the envelope of variation, we can study who uses these variables and why.

When analyzing the double modal in the Verilogue corpus, my colleagues and I looked at consultations in which the variable was present versus those where it was absent, and we analyzed the social factors of the speakers using the variable. In Hasty et al. (2012), we found that the presence of a double modal in a medical consultation was constrained by the length of time a doctor had been in practice, the doctor's gender, and the employment status of the patient. So, while we were unable to estimate true usage percentages, we were able to study how double modal use is governed by social factors and interpret what these social factors tell us about the double modal.

With hard to study variables like the double modal, constraints on usage may have more to do with pragmatics than is the case for other sociolinguistic variables. In our study, double modals were largely used during discussion of treatment. This context is characterized by asymmetric negotiations surrounding a patient's treatment plan. Doctors are expected to provide recommendations for the patient's care based on their expertise (e.g., *You'll need to get an MRI*). These recommendations generally take the form of directives (suggestions, orders, etc.), yet these directives pose a threat to a patient's negative face (Brown & Levinson, 1987) and as a result run the risk of not being followed. Doctors thus may have the goal of not only providing a patient with well-informed

health recommendations but also doing so without violating politeness constraints. Based on our results, we hypothesized that doctors with more experience (more years in practice) had learned to mitigate their directives through the use of double modals, which couch their directive in epistemic modality and thus lessen the force of the directive by making it appear more like a suggestion than a direct order (e.g., *you may should* rather than *you should*). This finding was later confirmed in a separate language attitude study (Hasty, 2015), which showed that a doctor using a double modal in a discussion of a treatment plan was judged as being more polite than the same doctor with the double modal removed.

Finding a corpus that included many instances of potentially face-threatening negotiations—the pragmatic situation most suitable for the double modal’s usage—was the key to finding the hitherto elusive double modal production data. Medical consultation corpora such as the Verilogue corpus may thus be quite useful for studying variables like the double modal, which have resisted traditional methods given their low occurrence. Additionally, even if we have to forego difficult (if not potentially impossible) attempts to delineate the envelope of variation for sociolinguistic features that have no clear co-variants, we can still find ways to identify important social constraints on usage. And since many hard to find and hard to study variables have resisted traditional sociolinguistic methodologies, we know very little about their constraining social and pragmatic factors, so any information we collect through studies of usage is a step in the right direction.

References

- Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Cheshire, J., Kerswill, P., & Williams, A. (2005). On the nonconvergence of syntactic and phonological variation. In P. Auer, F. Hinskens, & P. Kerswill (Eds.), *Dialect change: Convergence and divergence in European languages* (pp. 135–167). Cambridge: Cambridge University Press.
- Green, L. (2002). *African American English: A linguistic introduction*. Cambridge: Cambridge University Press.
- Hasty, J. D. (2014). We might should be thinking this way: Theoretical and methodological concerns in studies of syntactic variation. In R. Zanuttini & L. Horn (Eds.), *Micro-syntactic variation in North American English* (pp. 269–293). New York: Oxford University Press.
- Hasty, J. D. (2015). Well, he may could have sounded nicer: Perceptions of the double modal in doctor-patient interactions. *American Speech*, 90(3), 347–368.
- Hasty, J. D., Hesson, A., Wagner, S. E., & Lannon, R. (2012). Finding needles in the right haystack: Double modals in medical consultations. *University of Pennsylvania Working Papers in Linguistics*, 18(2), 41–47.
- Kozloff, J., & Barnett, J. (2006). *Verilogue corpus of patient-provider interaction*. Horsham, PA: Verilogue, Inc.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center For Applied Linguistics.
- Lavandera, B. (1978). Where does the sociolinguistic variable stop? *Language in Society*, 7, 171–182.
- Pichler, H. (2010). Methods in discourse variation analysis: Reflections on the way forward. *Journal of Sociolinguistics*, 14, 581–608.