English auxiliary contraction and the locus of variability

A common thread in work on sociolinguistic variation is that variable linguistic alternations are best modelled as probabilistic versions of invariant linguistic processes (Labov, 1969; Cedergren and Sankoff, 1974; Guy and Boberg, 1997). Work in this tradition has explicitly argued against treating variable alternations as the purview of a cognitive system distinct from the grammar; instead, variable and categorical phenomena are both understood to be represented grammar-internally (Weinreich et al., 1968). Motivating this proposal is the observation that variable and categorical phenomena tend to be conditioned by the same sorts of factors (Guy and Boberg, 1997; Coetzee and Pater, 2011). Localising variable and categorical phenomena in different domains would necessitate that the conditioning factors on each would have to be duplicated, resulting in unnecessary redundancy. The preferable alternative is to represent variation grammar-internally.

The present paper accepts this line of argumentation, but observes that not all conditions on variable phenomena lend themselves to a grammatical treatment in this way. This observation is based on data on contraction of the auxiliaries *is, has,* and *will* in the Switchboard (Godfrey et al., 1992), Fisher (Cieri et al., 2004), and Philadelphia Neighborhood (Labov and Rosenfelder, 2011) corpora. Auxiliaries were examined after full noun phrase (i.e., non-pronoun) subjects, and each auxiliary's subject was coded for length in words. Contraction of each auxiliary is strongly dispreferred as subjects increase in length (Figure 1).

Crucially, subject length differs from other conditions on variable phenomena because it is not found to condition categorical alternations in the world's languages: there is a common understanding that "grammars can't count". Expanding on the aforementioned proposal that shared conditions on categorical and variable phenomena motivate localising both in the grammar, we put forth the converse: when a variable phenomenon is conditioned by an effect which does **not** condition categorical alternations, that effect must **not** be localised in the grammar. The effect of subject length on contraction, we propose, is one such effect. We close with a discussion of what this subject length effect on contraction may be attributed to, if it is not grammatically-encoded. We propose that the effect may stem from constraints on production planning (e.g. Ferreira, 1991), hypothesising that longer subjects prevent a speaker from planning ahead to the auxiliary, which subsequently inhibits contraction.
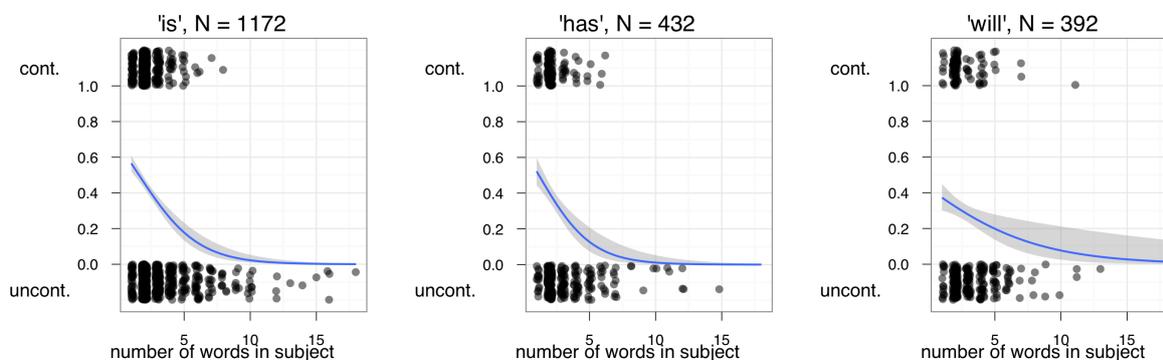

Figure 1: Effect of subject heaviness on contraction of three auxiliaries after non-pronoun subjects. Each point represents one token, coded as contracted (cont.) or not (uncont.). Values on the y-axis are for interpretation of the GLM smoothing line.