Take *that* away, and what do you get:
A study on complementizer variation in American English

Shelley Feuer (*The New School*), Matt Stuck (*New York University*)

This study uses multivariate analysis to investigate the effects of style shifting and linguistic constraints on the probabilistic distribution of the English complementizer, *that*. We conduct a side-by-side comparison of two different corpus-based sources which represent both formal and casual discourse styles. *That*-variation is described by variationists in terms of its stylistic and language-internal constraints on production (Kroch and Small, 1979; Rohdenburg, 1996; Torres and Walker, 2009). English complementizers vary between overt and null forms: "We know CP[that/Ø they're there because of their influence on other objects]"

To contrast complementizer variation across discourse registers, we compare data from lectures given by professors at Yale University (http://oyc.yale.edu/) with data from sociolinguistic interviews (drawn from the Buckeye corpus). In the Yale corpus, nearly 600 tokens were extracted from 11 lectures, including all tokens in contexts where a complementizer is licensed. In this formal register, the overall deletion rate was 39%. By comparison, the rate of complementizer absence in the sociolinguistic interviews, constituting an appreciably less formal discourse setting, exceeded 75% of all tokens (N>1200, drawn from 16 speakers), indicating that register has a significant effect on the speaker's omission of *that*. We contend that the notion of "grammatical ideology", proposed by Kroch and Small 1978, accounts for this disparity in complementizer expression between the two discourse types. K&S argue that prescriptive notions of linguistic correctness favor the use of explicit, overt grammatical structures in higher status speakers and more formal styles.

A binomial step-up/step-down analysis using GoldVarb shows that linguistic factors, like adjacency of the matrix and embedded verbs, grammatical subject, and the lexical identity of the main verb significantly affected the choice of the variant. These results suggest that the Complexity Principle also affects the use of *that*. This theory argues that "any elements capable of delaying the processing…of the overall sentence structure favor the use of an explicit signal of subordination" (Rohdenburg 1996, pp. 161). In our data, an increase in complexity of the verb form consistently leads to an increase in *that*, and increased syntactic complexity of the main clause (e.g., non-adjacency of verb and complementizer) also promotes *that*.

We also find significant effects of matrix verb etymology, which we argue to reflect both Rohdenburg's complexity principle, and Kroch and Small's notion of prescriptivism. Verbs of Latinate or French origin favored retention of *that*. Rohdenburg proposes that less frequent verbs (like most Latinate verbs in English) trigger an increased burden on the cognitive recognition of semantic properties, while Kroch and Small argue that Latinate verbs are perceived as more prestigious than frequently occurring Germanic verbs, triggering the more prescriptively correct structure.

Overall, our data show that a convergence of social and linguistic factors determines the distribution of complementizers in ways that reflect both a speaker's prescriptive beliefs and their capacity for internalizing a probabilistic syntactic grammar.

References

Cacoullos, R. and Walker. J. (2009). On the persistence of grammar in discourse formulas: A variationist study of *that*. *Linguistics, 47.1*: 1-43.

Kroch, A. and C. Small. (1978) Grammatical ideology and its effects on speech. *Linguistic Variation: Models and Methods.* Ed. David Sankoff. New York: Academic, 45-55.

Pitt, M., Johnson, K., Hume, E., Kiesling, S., & Raymond, R. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. Speech Communication, 45, 90-95

Rand, David & Sankoff, David (1990). GoldVarb: A variable rule application for the Macintosh. Montreal, Canada: Centre de recherches mathématiques, Université de Montréal.

Rohdenburg, G. (1996). Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics, 7*(2), 149-182. Retrieved from https://ezproxy.library.nyu.edu/