

What corpora can tell us about the link between grammatical variation and language complexity

Benedikt Szmrecsanyi (KU Leuven)

Abstract

In this presentation, I investigate the relative complexity (i.e., difficulty, see Miestamo 2009) incurred by having to choose between competing grammatical variants. While variational linguists provide overwhelming evidence for the existence, ubiquity, and systematicity of variable patterns — or “alternate ways of saying ‘the same’ thing” (Labov, 1972: 188), as in *Tom picked up the book* versus *Tom picked the book up* — there are still language mavens and theoretical linguists who dismiss or deplore variability as a matter of doctrine or explain it away (erroneously) as noise or negligible. Nonetheless, the assumption that grammatical variation *could* create undue complexity for language users is not entirely unreasonable. The idea that grammatical variation might burden language production deserves scrutiny not primarily because language users are forced to make grammatical choices — after all, using language *always* entails plenty of choice-making — but additionally because grammatical variation (as opposed to e.g., lexical variation) is typically conditioned probabilistically by any number of contextual constraints (constituent length, animacy, information status, etc.). Even before language users can make a choice as a function of the naturalness of a grammatical variant in a specific linguistic context, they need to check that linguistic context for the various constraints that regulate the variation at hand. It follows that this extra cognitive work must increase cognitive load. Or does it?

Against this backdrop, I report on a study that explores the link between production difficulty and grammatical variability using a corpus-based research design. The idea is that if isomorphism à la Haiman (1985) and No Synonymy à la Goldberg (1995) are design features of human languages, then variation — to the extent that it exists — should be suboptimal. Suboptimality, in turn, should be measurable by quantifying the extent to which variation contexts attract production difficulties.

Contrary to expectation, analysis based on a sub-sample of the Switchboard Corpus of American English (285 transcripts, 34 speakers) shows that the presence of variable contexts does not positively correlate with two metrics of production difficulty, namely filled pauses (*um* and *uh*) and unfilled pauses (speech planning time). When 20 morphosyntactic variables are considered collectively ($N = 6,268$), there is no positive effect. In other words, variable contexts do not correlate with measurable production difficulties. These results challenge the view that grammatical variability is somehow sub-optimal for speakers.

I will conclude by speculating that the putative difficulties introduced by optionality in syntactic structure or morphological realization are offset by a number of counterbalancing benefits of the flexibility to express the same grammatical concept using more than one form.

References

Goldberg, Adele E. 2003. Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences* 7(5). 219–224. <https://doi.org/10/fc25ng>.

Haiman, John. 1980. The Iconicity of Grammar: Isomorphism and Motivation. *Language* 56(3). 515.
<https://doi.org/10.2307/414448>.

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

Miestamo, Matti. 2009. Implicational hierarchies and grammatical complexity. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*, 80–97. Oxford: Oxford University Press.

