To be grammatical or not to be grammatical – is that the question? Evidence for gradience.

Chomsky (1955/1975) states that "grammar sets up a sharp division between a class G of grammatical sentences and a class G′ of ungrammatical sentences", and almost all subsequent theories of grammar build in such a sharp division. The acceptability judgments offered in support of these theories often include intermediate levels (marked with one or more "?"), but of the judgments published in Linguistic Inquiry 2001-2010, 94% are either starred or rated as fully acceptable, even though 81% occur in articles employing intermediate levels.

To check whether non-linguists perceive the divide between G and G′ as equally sharp, we randomly sampled 50 starred sentences ("*-items", $\in$ G′) and 50 unmarked sentences ("OK-items", $\in$ G), all of them from papers that employed a scale with three levels or more. We then had 80 non-linguists rate the 100 sentences, using a 7-point acceptability scale (1 "fully unnatural" to 7 "fully natural"). See Sprouse, et al (2013) for related work.

Participants' ratings correlate well with the LI judgments ($\rho_{pb} = .66$, $R^2 = .43$), rarely flatly contradicting them. Only 3 *-items received a mean rating > 5, and only 4 OK- items received a mean rating < 3. Yet, there are important discrepancies between the LI judgments and the experimental ratings.

The latter cover the whole space on the experimental 7-point scale rather than clustering at the two extremes. The intermediate range is densely populated, both for individual ratings and for mean ratings. In fact, 43 out of 100 items received mean ratings between 3 and 5. The graph shows a steady increase from 1 to 7, and no sign of the S-curve one would expect as a noisy approximation of the step-function corresponding to a sharp binary distinction.

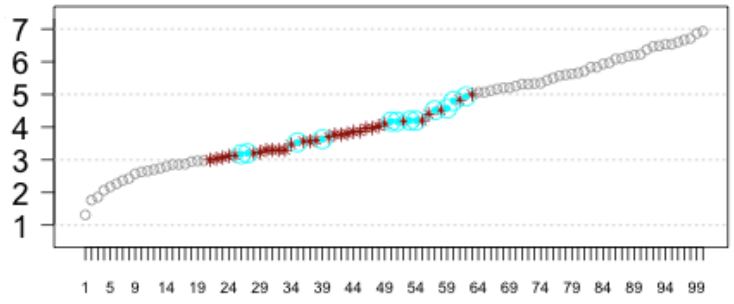

Figure 1: Mean ratings by the non-linguists (y-axis) for the 100 LI items (x-axis; in ascending order). Items with a mean rating of < 3 or > 5 are in gray. Items with a mean rating of ≥ 3 or ≤ 5 are in red (*-item) or blue (OK-item).

The observed gradience is not due to aggregation, as evidenced by the fact that individual ratings also cover the entire scale and cluster in the intermediate range. Known performance factors cannot account for the prevalence of intermediate values nor for the shape of the curve. The fact that most items with intermediate ratings were judged ungrammatical in LI also argues against attributing all gradience to performance, as suggested inter alia by Newmeyer (2003).

We conclude that gradience is not an epiphenomenon or artefact and should be built into theories of grammar.

– Chomsky, N. (1955/1975). *The logical structure of linguistic theory.* Chicago: University of Chicago Press

– Newmeyer, F. J. (2003). Grammar is grammar and usage is usage. *Language* 79, 682-707.

– Sprouse, J., Schütze, C. T., & Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001-2010. *Lingua* 134, 219-248.

(Word Count: 495)