

Main Study. Our main study uses attested sentences (sampled from Reddit comments) whose complementizers either match or mismatch a particular predicate’s biases according to its ROGATIVITY and RESPONSIVITY. For each of the 15 target predicates, we selected 10 sentences in which the verb selected for a declarative clause headed by *that*, and 10 where the verb selected an interrogative clause headed by *whether*. Attested examples of rogative and antirogative predicates selecting declarative and interrogative clauses, respectively, were not difficult to find in nearly every case. All selected sentences had at least one sentence of left context in the same comment.

Each selected sentence was used to create 3 items that varied in CONTENTFULNESS: (i) a morphosyntax-only item where all open class items were bleached while retaining the morphosyntactic structure of the item—e.g. (4a) (full sentence truncated was transformed into (4b)); (ii) a single sentence item, which is just the original sentence sampled from Reddit: e.g. (4a); and (iii) a multi-sentence item, which is that sentence with one additional sentence of left context.

- (4) a. However, I was hoping whether an upperclassman could help me [...].
b. I was hoping whether a certain thing could happen.

For each of these three items, we additionally make *baseline* items by flipping the complementizer. The idea behind creating these items is to provide a baseline for the improvement we should expect from increasing the amount of context a participant is given, irrespective of whether the sentence is attested (which the baseline variant is not, by construction). This process yielded 1,800 total items, which we partitioned into lists that only contained items from one of the three CONTENTFULNESS types. We recruited 1460 participants via MTurk to rate these sentences, with no participant seeing alternates of the same base sentence. Participants were prompted as in the norming study.

Analysis. To analyze the data, we compute the mean score for each item, then take the difference between each original item and each baseline item that corresponds to it, which yields a baseline correct IMPROVEMENT score. The mean IMPROVEMENT score for each verb with each complementizer and each CONTENTFULNESS level is plotted in Figure 2.

We fit a linear mixed-effects model to the IMPROVEMENT score in *brms*, with fixed effects for COMPLEMENTIZER (sum-coded), ROGATIVITY, CONTENTFULNESS (Helmert-coded with *morphosyntax only* as the reference), and all of their two- and three-way interaction, as well as by-verb and by-sentence random intercepts; by-verb random slopes for COMPLEMENTIZER, CONTENTFULNESS, and its two-way interaction; and by-sentence random slopes for complementizer. (Adding RESPONSIVITY to this model did not improve fit according to WAIC.) We find a reliably positive intercept and reliable main effects of ROGATIVITY and CONTENTFULNESS. No other effects are reliable in that their 95% credible intervals include 0.

The reliably positive intercept tells us that simply adding morphosyntax without any content improves all sentences across the board. This is consistent with theories that posit that a predicate’s acceptability with a particular type of subordinate clause should vary with morphosyntactic markers (e.g., Özyıldız, 2021). The reliable main effect of CONTENTFULNESS is driven by a positive effect of adding a single sentence’s worth of content. Adding additional context actually reliably (but to a lesser extent) degrades the improvement relative to the baselines (though it does not wipe it out). (In interpreting these results, it is important to remember that we are analyzing improvement relative to baseline, so this slight degradation from adding extra-sentential context does not imply that the multi-sentence variants themselves were rated worse than their single-sentence counterparts. Indeed, it turns out they were rated better; the baselines were simply also rated better.)

The main effect of ROGATIVITY is more surprising and suggests, based on analysis of the coefficients: the more rogative predicates are, the more reliably worse their improvement is across the board relative to antirogative predicates. This is surprising because it holds for both the *that* sentences and the *whether* sentences. Said another way: antirogatives improve with additional context to about the same extent regardless of whether they are found in a *that* or *whether* sentence. That antirogatives improve more with *whether* sentences than rogatives when additional context is provided is consistent with existing theories that posit that a predicate’s acceptability with a particular type of subordinate clause should vary across discourse contexts (e.g., Theiler et al., 2019). But no extant theory predicts that this effect should hold of both declaratives and interrogatives.