

Tell it like it (usually) is(n't): Speakers' mention of instruments in Reddit corpus
reflects instrument (a)typicality

Kapron-King, A., Fletcher, L., Tarighat, A., Dobрева, R.,
Droop, S., Cummins, C., & Rohde, H. [University of Edinburgh]

If speakers use language to describe their world, one would expect to encounter language that describes the kinds of situations in which speakers are likely to find themselves, i.e., situations that are frequent and typical in the world (TELL IT LIKE IT IS). Prior work indeed affirms a role for real-world knowledge in people's expectations about what a speaker is going to say [1,2]. On the other hand, if one of the things that speakers do with language is to convey non-inferable information about particular situations, one would expect language that favors newsworthy content, particularly if speakers make rational decisions about the inclusion of syntactically optional elements (TELL IT LIKE IT ISN'T). Prior work on speakers' productions confirms a preference for mentioning the atypical: Event descriptions are more likely to include the instrument when it is atypical (e.g., an icepick rather than a knife for a stabbing [3]) and object descriptions are more likely to include a modifier when the property is atypical (e.g., blue rather than red strawberry [4]; wool rather than ceramic bowl [5]). Such work has primarily tested speakers' productions in psycholinguistic experiments. Here we use naturally occurring productions (extracted from social network forum www.reddit.com). We target speakers' choice to talk about the presence of an instrument ("with") or its absence ("without"), see examples in Table 1. In order to compare the TELL IT LIKE IT IS versus -ISN'T accounts, we extract verb-object-instrument triplets that appear in Reddit comments. We test how commenters' choice to mention the presence or absence of an instrument varies with the typicality of the instrument (as indicated by a set of naïve participants). A rational strategy for instrument mention would predict "with" mentions for atypical instruments and "without" mentions for typical ones (TELLING IT LIKE IT ISN'T).

Methods. We automatically extracted mentions of instruments from 4 months of Reddit comments with templates for positive "with" and negative "without" instances, removing obscenities, brand names, and non-concrete nouns, yielding 9,621 cases. A manual filtering removed context-dependent instruments (*same water*) and cases where the noun was not enabling the action (*eat burger with cheese*), yielding 499 verb-object-instrument triplets. These triplets were presented in batches of 49-50 to Prolific participants (N=206, £1.25 payment) who gave typicality ratings on a scale of 1 to 11 (Fig 1). The resulting data consisted of 9826 typicality ratings after exclusions.

Results. Under the TELL IT LIKE IT IS account, we'd expect speakers to mention the presence of typical instruments and the absence of atypical ones. Instead, instruments in "with" mentions received only mid-range typicality ratings (mean rating 6.6) and those in "without" mentions received higher typicality ratings (mean 7.4). Fig 2 shows individual instruments' proportion of positive "with" mentions (out of all "with" and "without" mentions of that instrument). As can be seen, "with" mentions decrease as typicality increases. We used a logistic regression to model the binary outcome of each Reddit comment's polarity ("with" versus "without") with a fixed effect for typicality and a random effect for instrument. The results show a main effect of typicality (β : -0.180, SE=0.0707, $p=0.01$). Table 1 shows the pattern for the instrument *spoon* and an illustrative set of triplets about *teeth brushing*.

Our findings suggest speakers make rational choices to talk about uninferable aspects of a situation, i.e., the presence of atypical instruments or the absence of typical ones. We thus extend prior work on rational production, showing a pressure for the inclusion of informative content in the mention of instruments in naturally occurring language.

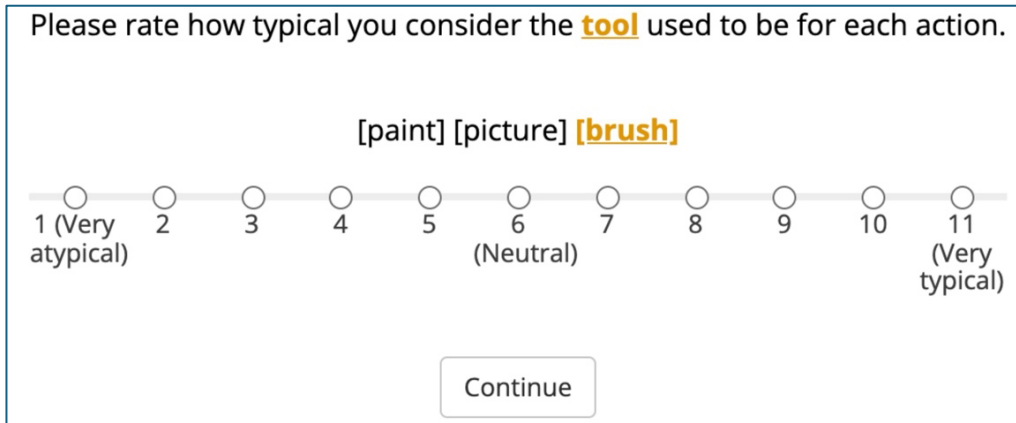


Figure 1: Task setup for elicitation of instrument typicality

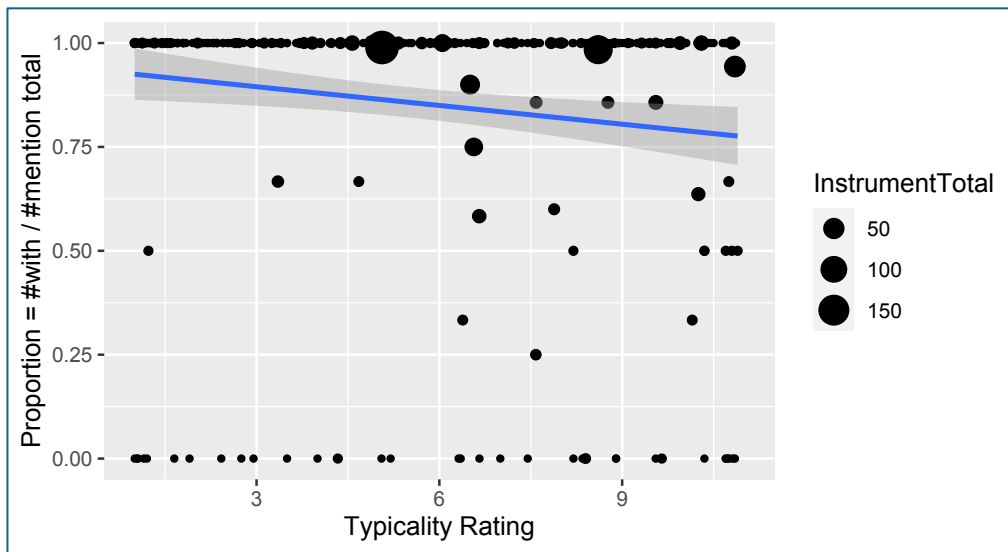


Figure 2: Mention of instruments using “with” vs “without” in Reddit corpus

| verb-object-instrument triplet | Reddit probability | Typicality rating |
|--------------------------------|--------------------|-------------------|
| eat chili spoon | p(“with”) = 0.0 | 9.6 |
| eat spaghetti spoon | p(“with”) = 1.0 | 3.0 |
| brush teeth toothpaste | p(“with”) = 0.4 | 10.3 |
| brush teeth fingers | p(“with”) = 1.0 | 3.4 |
| brush teeth wrong hand | p(“with”) = 1.0 | 3.3 |
| brush teeth nail file | p(“with”) = 1.0 | 1.1 |
| brush teeth toilet brush | p(“with”) = 1.0 | 1.0 |
| brush teeth 12 gauge shotgun | p(“with”) = 1.0 | 1.0 |
| brush teeth brick | p(“with”) = 1.0 | 1.0 |
| brush teeth hairbrush | p(“with”) = 1.0 | 1.0 |

Table 1: Sample of Reddit usage (probability of “with”) and mean typicality rating (1-11 scale)

[1] Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1). 133–156. [2] Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207. 203–205. [3] Brown, P. M. & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology* 19. 441–472. [4] Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research* 32. 3–23. [5] Mitchell, M., Reiter, R. & Van Deemter, K. (2013). Typicality and object reference. *Proceedings of the Annual Meeting of the Cognitive Science Society* 35. 3062–3067.