The impact of the wording of tasks in linguistics experiments and violations of Grice's second maxim of quality – Follow Up

Alexandre Cremers, Lea Fricke, Edgar Onea

We conducted a follow-up study of our experiment *The impact of the wording of tasks in linguistic experiments and violations of Grice's second maxim of quality*. The experiment had two goals. First, it investigated how violations of Grice's (1975) second sub-maxim of quality are perceived. The maxim of quality is shown in (1)

(1)   Try to make your contribution one that is true.
          a. Do not say what you believe is false.
          b. Do not say that for which you lack adequate evidence.

We wanted to find out how participants react to violations in different contexts, which vary with respect to the probability that a statement uttered by a person x is true. On one end of the spectrum we employed, was a situation in which the speaker utters a wrong statement while her evidence was such that she could be a 100 percent certain that the statement is false; and on the other end of the spectrum there was a situation in which the speaker utters a true statement having 100 percent certainty that it is true. In between these extremes, lie situations in which the statement was either true or false, with varying probability that this is the case given the speaker's knowledge.

The second objective of the experiment was methodological. We tested how different prompts affect the responses of participants. In particular, it was tested whether a given prompt is sensitive to the truth of the statement or the evidence the speaker had who made the statement. We tested the prompts listed in (2) in the context of a short story.

(2)   a. Is x's statement true?

          b. Was the answer correct?

          c. Is x right?

          d. Was x's statement natural in this context?

          e. Was x justifed in saying that?

          f. Based on her behavior in this situation, do you consider x trustworthy?

A further methodological aspect that was studied was the effect of the response task. To that end, we compared judgments collected with binary response options, meaning participants had to choose between the answers *Yes* and *No*, with judgements made on a Likert scale from 1 to 7, with the endpoints labelled as *Definitely yes* and *Definitely no*.

The data we obtained in the first run of the experiment overall confirmed our hypotheses that different prompts vary regarding their sensitivity to the probability that the statement is true. We found that phrasings which explicitly target truth-value judgements are less susceptible to variations of the probability than wordings targeting the appropriateness in the context. We also found that the response data collected with a Likert scale showed less noise than the data collected with binary response options. However, an unresolved issue remained. The data for some prompts showed a pattern that can be interpreted to indicate that some participants misunderstood an important aspect about the context. Therefore, in the follow-up experiment we changed the context description to make it more explicit. We also added an additional

comprehension question that tested whether the aspect in question was understood correctly. Moreover, we increased the number of participants in order to reduce noise in the response data.

The follow-up study yielded the following results. For the prompts (2a-c) the truth of the utterance is judged independent of the evidence. We conclude that these prompts are well-suited to be used in truth-value judgement experiments. The other prompts (2d-e), were sensitive to both truth and evidence. (2e) seems to require strong evidence as well as truth, whereas (2d) and (2f) show a graded effect of evidence and only a moderate effect of truth. This group of prompts is suitable for experiments investigating different aspects of pragmatic communication. With regard to the response task, we found that, the Likert scale achieves the same statistical power as the binary response options with only half the sample size. Thus, we conclude that it is generally a better choice to use a Likert scale in an experiment than the binary response options.

We regard prompt (2d) as the test case for the force of the maxim of quality, as the question whether a statement is natural relatively neutrally inquires whether it appears likely that a person would make that statement in a given context. Interestingly, acceptance for true statements increased gradually with rising evidence, which means that a number of participants found it natural to utter a statement for which the evidence was less than certain. Thus, some people value truth over evidence, possibly due to the fact that even accidental truth may lead to success.