Scenes in reference generation: Preregistration Document

Philipp Stein and Dale J. Barr

February 22, 2018

GOAL: Test whether the generation of referring expressions depends on remembered aspects of a scene.

OVERVIEW: Participants view photos of real objects embedded in scenes, and are asked to generate a referring expression for one of the objects (the target). In each photo, the target appears in the presence or absence of a contrasting "competitor" object of the same category. An example is given in Figure 1. In this example, due to the presence of the competitor (the bitten apple), the speaker would need to refer to the target with a modifier, e.g., as "the apple without the bite." For each type of target, speakers would entrain on a description over three to five "training" trials. Then the same target would appear in a single test trial where the context would change such that the description used over the training trials would no longer be optimal. If the target appeared with a competitor at training, then the competitor would replaced by a visually-similar noncompetitor object from a different category (e.g., a pear; thus, in this example, "the apple" would be sufficient). Conversely, if the target appeared with a noncompetitor at training, then it would appear with a competitor at test. The key DV is the **misspecification** rate at test, which measures the extent to which during the test trials speakers rely on remembered descriptions from training.

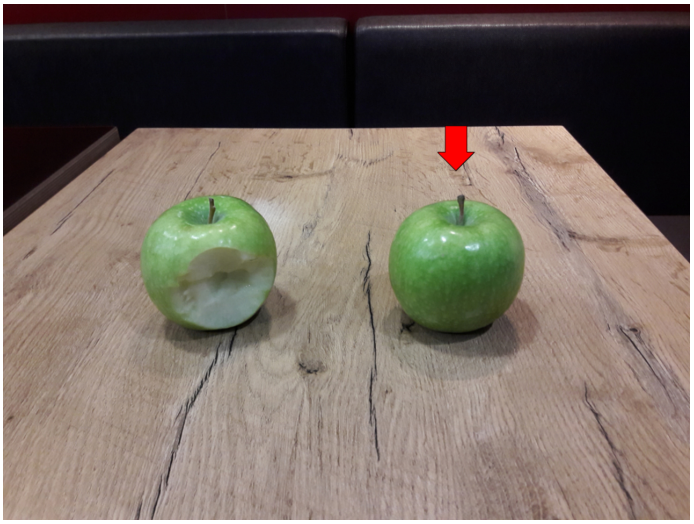

Figure 1: *An example training image with the target marked by a red arrow. A speaker would need a modifier to refer to this object, e.g., "the apple without a bite".*

Participants entrained on descriptions for targets in one scene, and the test trial either showed the target in the same scene, or in a different scene.

MAIN HYPOTHESIS: At test, speakers will rely more on remembered descriptions when the scene matches the training scene than when it mismatches, leading to a higher misspecification rate. This is a directional hypothesis, so we are pre-registering a one-tailed test.

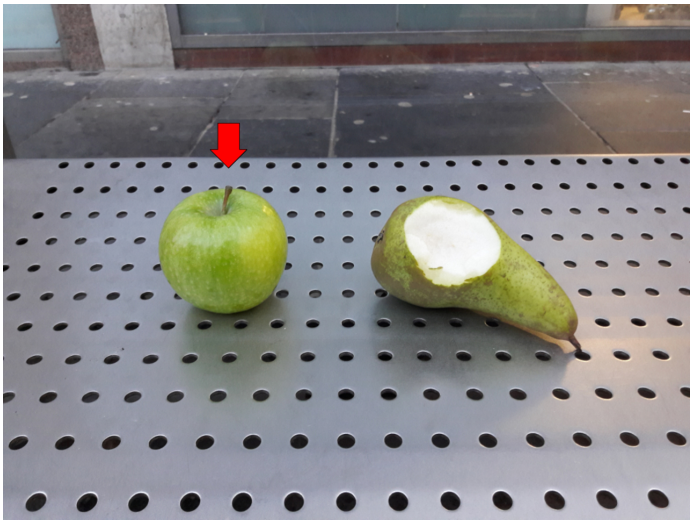

Figure 2: *An example test image with the target marked by a red arrow. A speaker would no longer need a modifier to refer to this object, e.g., "the apple" would be sufficient. Note also that the scene is incongruent with the training image.*

Participants

Forty-eight University of Glasgow Undergraduates will participate in our study. All subjects will be native English speakers. Subjects will provide written consent prior to the experiment and will be fully debriefed afterward. Ethics comply with the BPS Code of Conduct.

This study is part of a student project (Honors Thesis), and we anticipate being unable to complete collection of a full 48 participants before the assignment deadline. Thus, an analysis will be performed with a subset of the full data in order to meet the deadline, but the results of this analysis will not inform any subsequent analytic decisions. The results will not be considered definitive until the full set of 48 participants has been collected.

Experimental Setup and Task

The experiment will consist of the participant playing the role of the 'Director'. They will be informed that they are recording directions for a later participant who will play the role of the 'Matcher'. The Director will be informed that the Matcher will see the same objects in the same location, but that the objects may not be in the same location or orientation to the image they will see. During the recording, the experimenter will play the role of the Matcher and click on the object described on a display that appears behind the participant.

Design

The study will have two factors. The first factor is Shift Direction (Singleton-To-Contrast versus Contrast-To-Singleton), which represents whether a competitor was present at training and absent at test (Contrast-To-Singleton) or vice versa. The second factor is Scene Congruency (Congruent and Incongruent), forming a 2x2 within-participants design.

Materials

The materials comprise 40 sets of images, each set containing four photos: target with competitor in Scene A, target with foil in Scene A, target with competitor in Scene B, target with foil in Scene B. All of the images are freely available in the `resource/` folder of the github archive associated with this pre-registration.

Additionally, the tables in the associated database (`Philipp_Maxi.db`) contain the entire structure of the experiment, with `Session` associated each session number with a stimulus list (`ListID`), and the order of trials for each list given by the `ListOrder` table. The images displayed for each trial can be accessed in the `Resource` table.

Analysis

Our analysis will focus on two main categories of measurement: (1) speech content, use of modifier and fluency, (2) speech onset latency, the time taken to produce the to produce the first content word as measured from the beginning of the trial.

For each of the 44 sequences for each Director, we will transcribe and code audio from two recordings: (1) the final training trial and (2) the test trial. The final training trial will be necessary for to provide baseline data for the speech onset latency in the test trial. Each trial will be transcribed and coded for fluency and adjective use. Fluency will be coded into one of four categories, as shown in the table below.

Code	Description	Example
FL	Fluent Speech	"The clean plate", "The plate"
UP	Unfilled Pause (After speech onset)	"The ... uh ... clean plate"
FP	Filled pause (um/uh)	"uh ... clean plate"
RE	Repaired phrase	"Plate ... yea ... Plate", "Plate ... clean plate"
LS	Lengthened Speech	"The cle(eeea...)n Plate"

We will also code whether or not a modifier is used, defined by the following categories:

Code	Description	Example
NO	No modifier	"The plate"
PR	Pre-nominal modifier	"The clean plate"
PO	Post-nominal modifier	"The plate the clean one"
DE	Deleted adjective	"clea ... no the plate"
AS	Addition due to self-repair	"Plate ... clean plate"

Onset times of utterances will be identified and entered into a data table in milliseconds. These criteria will be applied when identifying utterance onsets:

1. Trials will be deleted if speech was unidentifiable.
2. Any filled pauses or articles will be ignored; speech onset will be identified as the first content word.
3. If Directors correct themselves after an error onset of the correction will be recorded, however repaired utterances will not be used for the analysis of speech onset.

Data preparation

Directors may opt to always use a modifier regardless of whether presented with a competitor or noncompetitor. This presents a problem differentiating a modified response from habitual use of modified responses or speakers "hyperdescribing" targets. We will look at the final training trial for all cases in which the target

appeared with a noncompetitor. For any cases where speakers used a description that would distinguish the target from the (non-present) competitor, we will remove the corresponding test trial from the analysis. We will also remove the complete data from any participants for whom more than half of the trials would be deleted based on this criterion. We will do the same for targets: we will remove the full data for any targets for which half of the test trials would be deleted by this criterion.

Analysis and Predictions

The statistical analysis for the production data will be performed using a linear mixed effect model with directors (Subjects) as a random factor. All analyses will use maximal random effects structure justified by the design. This implies:

- by-subject random intercepts and by-subject random slopes for both factors and their interaction;
- by-item random intercepts and by-item random slopes for both factors and their interaction.

We will derive p-values using the t-to-z heuristic (deriving the p-values using the standard normal distribution for the t-statistic), because this allows us to perform our critical one-tailed test. Models will be estimated using the lme4 package in R (Version 1.1-12 or higher).

Our main prediction is that speakers will be more likely to misspecify the target (use unnecessary modifiers or leave out modifiers) in the Congruent condition compared to the Incongruent condition; in other words, we predict a main effect of Scene Congruency. For this analysis, we will use a generalized linear mixed model with a logit link and binomially distributed error variance. We will use a one-tailed with the alpha level for this test set at .05.

Our second main prediction concerns speech onset latency. We predict that speakers will experience more difficulty shifting from the trained description to a contextually appropriate description when the test scene is congruent with the training scene. This analysis will exclude any trials where the target was misspecified.

Parameters will be estimated under maximum likelihood (REML=FALSE) using a linear mixed effect model with identity link and gaussian variance. The dependent variable is the speech latency for the test trial minus the speech latency for the training trail in that sequence, in other words, the change in speech latency by abandoning the entrained description. This anlaysis will be two-tailed with alpha set to .05.