Perceptual Difficulty Differences Predict Asymmetry in Overmodification with Color and Material Adjectives

Leyla Kursat & Judith Degen [*]

Abstract. [lk: last thing to change: this is taken from the abstract]. When referring to objects, speakers are often more specific than they need to be for establishing unique reference (e.g., "the green plastic bottle" instead of "the plastic bottle" in Fig. 1). Adjectival overspecification patterns are not random, but structured: color adjectives are produced redundantly more often than size or material adjectives [1-5]; color adjectives are more likely to be produced redundantly with increasing scene variation [5-7]; and adjectives are more likely to be produced redundantly, the more atypical the property denoted by the adjective is for the object under discussion 5;8-9]. The only current computational model of referring expression production that accounts jointly for all of these patterns is couched within the Rational Speech Act framework [9] and assumes that adjectives differ in how noisy, and consequently, how useful they are for the purpose of establishing reference [5]. One hypothesis about the nature of this noise is that it reflects the perceptual difficulty of establishing whether the property denoted by the adjective holds of the contextually relevant objects. Here, we take a first step towards testing the prediction that systematic differences in the overmodification patterns observed for color and material adjectives can be explained by a difference in perceptual difficulty of establishing whether objects are of a particular color or material. In Exp.1, we norm the perceptual difficulty associated with establishing whether an object exhibits a color or material and select objects with highest and lowest perceptual difficulty for testing in Exp. 2. In Exp. 2, we test in a reference game wheter adjectives that denote more perceptually difficult properties indeed are less frequently produced redundantly.

Keywords. reference; perception; overinformativeness; experimental pragmatics

1. Introduction. [lk: what are overinformative referring expressions, the systematicities with which redundant adjectives are produced]
[lk: degen et al.'s computational model of referring expression production]
[lk: different hypothesis about the source of the adjectival noise]
[lk: overview of this study]

2. Experiment 1: Measuring perceptual difficulty.
First, we need to norm the perceptual difficulty associated with establishing whether an object exhibits a color or material. Through a timed forced choice task we collected norms for color and material properties of 81 images.

2.1. PARTICIPANTS. We recruited 120 participants through Amazon Mechanical Turk. We excluded participants who were self-reported non-native English speakers (n=4) and participants with accuracy lower than 75% (n=11).

---

[*] Authors: Leyla Kursat, Stanford University (lkursat@stanford.edu) & Judith Degen, Stanford University (jdegen@stanford.edu).

2.2. PROCEDURE. Participants saw images of objects with color or material adjectives and were asked to indicate whether the object had the property denoted by the adjective or not. Their task was to indicate "yes" or "no" by pressing the F or J key as quickly as possible. If participants did not respond within 4 seconds, the trial timed out. When participants responded correctly, a green border appeared around their selection, and when they responded incorrectly a red border appeared.

We collected perceptual difficulty norms for 12 objects, that each occurred in two or three different materials and in three different colors. All resulting 81 images were separately normed for object nameability, feature nameability, object typicality and feature typicality. Every participant saw each image once and we collected 30 judgements for each image with correct and incorrect color and material adjectives. Incorrect color and material adjectives were randomly generated online for each participant and were selected from adjectives denoting properties of other images in the experiment.

2.3. RESULTS. Material words resulted in higher error rates (β= 0.40, SE=0.09, p<.0001) and greater response times (β=5.46, SE=4.73, t=11.55, p<.0001) than color words. We grouped the 8 image-material adjective pairs with the highest error rate and response times into a *high difficulty* group, and the 8 image-color adjective pairs with the lowest error rate and response times into a *low difficulty group*. [lk: more about grouping and selection of items, limitations of the task?]

3. Experiment 2: Production of referring expressions. The goal of Exp. 2 was to elicit production probabilities of redundantly mentioning color and material adjectives for the high- and low-difficulty items normed in Exp. 1. In a free production interactive reference game we tested whether adjectives that denote more perceptually difficult properties are less frequently produced redundantly.[1]

3.1. PARTICIPANTS. We recruited 100 participants through Amazon Mechanical Turk and randomly paired them into speaker-listener dyads to play a real time communication game (50 pairs) [lk: reference to Hawkins 205?]. We excluded games where participants reported a native language different from English.

3.2. PROCEDURE. On each trial, participants saw a display with 4 images and chat box. Both the speaker and listener saw the same images in different positions. One of the images was designated as the target image, and marked by a green border in the speaker's display. The speaker's task was to describe this target image to the listener using the chat box to send messages. The listener's task was to guess the target image by clicking. After the listener made a selection, both participants received feedback about whether the target image was selected and advance to the next trial.

Participants completed 32 trials. Of these, half were critical trials and half were filler trials. On critical trials [lk: add figure+ reference], the 4 images were of the same object and either *color* or *material* was redundant for distinguishing the target. On 8 high-difficulty trials, mentioning the material was redundant; on 8 low-difficulty trials, color was redundant. On filler trials, the 4 images were of different objects and both color and material mention were redundant for unique reference.

---

[1] Procedure, materials, analysis and exclusions were preregistered at https://osf.io/57c6u.

We classified the produced utterances as 'color-and-material' (redundant), 'only-color' or 'only-material'.

3.3. RESULTS. [lk: replicate the asymmetry?, given the norms - support for hypothesis, manipulate perceptual difficulty within property type..]

4. Experiment 3: Perceptual difficulty in context.

4.1. PARTICIPANTS.

4.2. PROCEDURE.

4.3. RESULTS.

5. General discussion.

6. Citations and references. Use author-date notation for in-text citations, for example: ... as noted recently by Jameson (2012) and Mateus (2014), drawing on insights from various other researchers (e.g., Nelson 1986:223–28, Martin 2003, Wellington & Johnson 2016), there have been numerous technological advances in the procedures used to publish research articles online. Include URLs or DOIs with active hyperlinks in citations. The *Semantics and Pragmatics* stylesheet sp.bst meets the LSA formatting requirements for the list of references, and so may be used. See info.semprag.org.