

# Beyond typicality: Lexical category affects the use and processing of color words

Madeleine Long (M.Long@ed.ac.uk)

Psychology Department, University of Edinburgh, Edinburgh, Great Britain

Paula Rubio-Fernandez (Paula.RubioFernandez@mpi.nl)

Multimodal Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

## Abstract

Speakers and listeners show an informativity bias in the use and interpretation of color modifiers. For example, speakers use color more often when referring to objects that vary in color than to objects with a prototypical color. Likewise, listeners look away from objects with prototypical colors upon hearing that color mentioned. Here we test whether speakers and listeners account for another factor related to informativity: the strength of the association between lexical categories and color. Our results demonstrate that speakers and listeners' choices are indeed influenced by this factor; as such, it should be integrated into current pragmatic theories of informativity and computational models of color reference.

**Keywords:** pragmatics, language production, language comprehension, reference, informativity.

## Introduction

The use of color adjectives is ubiquitous in everyday communication, and more frequent than other types of adjectives. In a classic study in experimental pragmatics, Sedivy (2004) observed that English speakers used scalar, material and color adjectives *contrastively* (e.g., they referred to 'the short pencil' when there was a longer pencil in the display), but they also used color adjectives *descriptively* around half the time (e.g., they referred to 'the pink comb' when that was the only comb in the display). Given the relatively high frequency with which people used color adjectives compared to scalar and material adjectives, Sedivy (2003, 2004) argued that color is part of the 'default description of objects.'

However, contrary to the default view of color, recent referential communication studies have shown that speakers' tendency to use color adjectives descriptively varies greatly depending on a number of factors, including the discriminability of the referent in the display (e.g., whether the display is monochrome or polychrome; Koolen et al., 2013; Long et al., 2020, 2021; Rubio-Fernandez, 2021), the density of the display (e.g., whether it is sparse or dense with objects; Clarke et al., 2013; Gatt et al., 2017; Rubio-Fernandez, 2019), and the position of the adjective (i.e., whether it is prenominal (e.g., 'the white house' in English), or postnominal (e.g., 'la casa blanca' in Spanish) (Rubio-Fernandez, 2016, 2019; Wu & Gibson, 2021).

The variability observed in these referential communication studies has been interpreted as an effect of the varying *efficiency* of descriptive color modification in different visual contexts, since not only contrastive but also

descriptive color adjectives can facilitate the listener's visual search for the referent, as attested in various eye-tracking studies (Tourouri et al., 2019; Rehrig et al., 2021; Rubio-Fernandez, 2021; Rubio-Fernandez et al., 2021).

The present study investigated one of the factors that have been shown to affect descriptive color use; namely, the lexical category of the noun. Previous studies investigating lexical effects have focused on *color typicality* (e.g., in fruits and vegetables with predictable colors such as bananas and tomatoes; Sedivy, 2003, 2004; Westerbeek et al., 2015; Rubio-Fernandez, 2016; Kreiss & Degen, 2020; Long et al., 2021). Here we wanted to look at lexical effects beyond color typicality and investigated how the *strength of association between lexical categories and color* (e.g., clothes vs appliances) affects the use and comprehension of descriptive color adjectives.

## Lexical effects beyond color typicality

Previous referential communication studies have shown that speakers use color more often when referring to objects that vary in color than objects with a prototypical color (e.g., yellow notebook vs yellow banana) as it is more informative (Sedivy, 2003, 2004; Westerbeek et al., 2015; Rubio-Fernandez, 2016; Kreiss & Degen, 2020; Long et al., 2021). Interestingly, speakers are not the only ones who make use of this informativity bias. Indeed, eye-tracking work has confirmed that listeners display the same bias when making predictions about upcoming content. For example, Sedivy (2003, 2004) observed that listeners expect prototypical colors to be used contrastively (e.g., when 'yellow' disambiguates between a yellow and a green banana) and non-prototypical colors to be used descriptively (e.g., 'the yellow notebook' to refer to a single notebook).

Going one step further, Rohde and Rubio-Fernandez (2022) recently showed that this bias extends to contexts where contrastive inferences are unnecessary (i.e., where objects need not be disambiguated by their color). Two competing hypotheses were tested in this study: the World Knowledge Hypothesis and the Informativity Hypothesis. The former posits that language production and comprehension are grounded in real world plausibility, thus situations that are frequent in the world should give rise to frequent utterances about those situations. This is represented in the following formula:

$$p(\text{referent} = \text{object}_i \mid \text{'yellow'}) \propto p(\text{object}_i, \text{color} = \text{yellow})$$

Here the listener's expectation that the yellow object being referred to is a banana would be higher than their expectation that the object is a strawberry, since encountering a yellow banana is a more common occurrence in the real world.

Conversely, the Informativity Hypothesis posits that listeners not only use knowledge about the real world to inform their expectations of upcoming content, but also pragmatic knowledge about speakers' use of language. This is represented in the following formula:

$$p(\text{referent} = \text{object}_i \mid \text{'yellow'}) \propto p(\text{object}_{i,\text{color}=\text{yellow}}) * p(\text{'yellow'} \mid \text{object}_{i,\text{color}=\text{yellow}})$$

Here, a listener that sees both a yellow banana and a yellow notebook and hears the speaker mention 'yellow' should anticipate that the speaker is talking about the notebook. This reflects the listener's pragmatic knowledge: even though bananas are more likely to be yellow than notebooks, speakers almost never describe bananas as yellow (unless their color is contrastive with another banana).

Findings from this study offered support to the Informativity Hypothesis, as listeners were found to look away from objects with prototypical colors upon hearing that color mentioned (e.g., participants looked away from a picture with two bananas and towards a picture with two shirts when hearing 'yellow' because the bananas' color is less likely to be mentioned as an identifying feature than the shirts' color). This suggests that listeners *expect* speakers to use color adjectives informatively, both in contexts that require deriving a contrastive inference (Sedivy, 2003, 2004; Kreiss & Degen, 2020) and in contexts where color is used descriptively (Rohde & Rubio-Fernandez, 2022).

The results of the above studies suggest that both speakers and listeners are sensitive to color typicality, omitting prototypical colors in reference production and expecting them to be omitted in reference comprehension. This sensitivity has been further documented in referential communication studies with objects of *atypical colors* (e.g., a blue banana), which are often mentioned descriptively (Westerbeek et al., 2015; Rubio-Fernandez, 2016; Kreiss & Degen, 2020; Long et al., 2021).

However, lexical effects in the production of descriptive color adjectives have been observed beyond color typicality: Even when objects lack a prototypical color, speakers tend to describe their color more often when color is a central property of the lexical category, than when it is not (e.g., color is more strongly associated with clothes than with geometrical shapes). This was evidenced in recent referential communication studies: when presented with monochrome displays of clothes, participants in Rubio-Fernandez (2016) produced color-modified descriptions like 'the yellow shirt' around 40% of the time, whereas participants in Rubio-Fernandez (2019) never produced 'the yellow triangle' in a monochrome display of shapes, and did so 40% of the time in polychrome displays.

The results of Rubio-Fernandez (2016) seem to challenge the view that descriptive color modification aims to facilitate

the listener's visual search for the referent (Rubio-Fernandez et al., 2021), since color is a useless visual cue in monochrome displays. However, when interpreted against the results of Rubio-Fernandez (2019), the earlier findings suggest that the tendency to mention the color of clothes is strong enough to override efficiency considerations—whereas that is not the case for geometrical shapes, which are not so strongly associated to color. The aim of the current study was to therefore investigate how the strength of the association between lexical categories and color affects the use of descriptive color modification in referential communication, above and beyond color typicality.

## Current study

Building on the above work, we tested whether speakers (Experiment 1) and listeners (Experiment 2) are sensitive to how strongly lexical categories are associated with color. We treated fruits and vegetables with predictable colors as our typicality baseline, and clothes as a lexical category with great color variability (i.e. without prototypical colors) but whose association to color is particularly strong. The aim of the study was to identify other lexical categories that varied in color but were not so strongly associated to color, and test speakers' production and listeners' expectation of color modification for members of those categories. For example, color seems intuitively more central to clothes than to appliances, but it is an open empirical question whether the differential strength of their color association is reflected in speakers' descriptive use of color adjectives to refer to these two categories, as well as in listeners' expectations of color modification.

Our main prediction was that speakers would use color adjectives more frequently for categories with stronger color associations and that correspondingly, listeners' judgments of upcoming content would reflect this informativity bias. For example, we predicted that speakers would be more likely to refer to 'the green dress' than 'the green hair dryer' in a display of singleton objects. Similarly, on the comprehension end, we predicted that after hearing the word 'green', listeners would anticipate the next word to be 'dress' over 'hair dryer.'

In Experiment 1, we tested the production side of our hypothesis through a web-based referential communication task in which participants were asked to identify targets for a hypothetical listener. In addition to the Fruit baseline (typical colors) and Clothes condition (strongest color association), we selected other lexical categories whose members vary in color (e.g., appliances) but for which color is not so central. It must be noted that these intermediate categories were selected based on intuition, rather than on an empirical basis.

In Experiment 2, we tested the comprehension side of our hypothesis through a web-based sentence completion task. The aim of this task was twofold: first, we wanted to investigate whether listeners would show the same informativity bias as speakers in Experiment 1. Second, we wanted to assess whether this informativity bias would extend beyond color typicality.

## Experiment 1

### Methods

#### Participants

A total of 323 native English speakers over the age of 18 were recruited through Prolific (a crowdsourcing platform) to take part in the study. All participants reported normal or corrected-to-normal vision and informed consent was obtained prior to testing.

#### Materials and Procedure

Participants were presented with 20 four-object displays. To increase their variability, 10 displays were monochrome and 10 were polychrome (see Fig. 1). The target objects (and monochrome displays) were the following colors: black, blue, brown, gray, green, orange, purple, red, yellow and white. The position of the target was counterbalanced across trials and marked with an asterisk.

The objects in the displays varied across five lexical categories: Appliances, Cars, Clothes, Fruits and Toys. Fruits was the only category with predictable colors. We intuitively ranked the remaining four categories according to the strength of their color association (in increasing order): Appliances, Toys, Cars and Clothes.

To avoid carry over effects across categories, the visual materials were distributed in five separate lists by lexical category. Participants were randomly assigned to one of these five lists and instructed to identify the target objects for a hypothetical addressee by typing their description in an open text box. Trials were randomized individually.



Figure 1: Sample displays from Experiment 1 for the lexical categories Fruits (top), Appliances (middle) and Toys (bottom).

All four objects were different in each display, rendering color redundant. As the only category with predictable colors, Fruits served as our baseline.

We predicted that Appliances (the lexical category that we assumed had the weakest color association) would not elicit descriptive color modification, similar to the color-predictable Fruits baseline. Conversely, we predicted that the other three lexical categories (which we thought had stronger color associations than Appliances) would elicit higher rates of descriptive color adjectives than the Fruits baseline.

### Results

The data were analyzed with the lme4 package (Bates et al., 2015) from R (R Core Team, 2021). Using logistic mixed effects regression, we modelled the binary outcome variable of Color Modification (1=Modification, 0=No modification) with Lexical Category as the fixed effect, Fruit as the reference level, and the maximal random effect structure by Participants and Items (Barr et al., 2013).

Confirming our predictions, the model revealed greater descriptive color modification for Cars, Clothes, and Toys than for Fruits (all  $p$ 's < .001; for the full model output, see Table 1), while the rate for Appliances did not differ from the baseline ( $p = .134$ ; see Fig. 2).

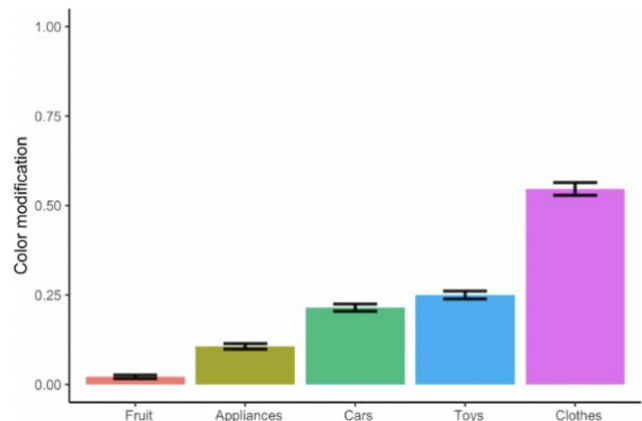


Figure 2: Average rates of descriptive color modification for the five lexical categories in Experiment 1.

Table 1: Model output from Experiment 1

Fixed Effect	Coefficient	SE	P-value
Appliances	1.646	1.098	.134
Cars	4.804	1.230	<.001
Clothes	10.859	1.437	<.001
Toys	5.382	1.223	<.001

Interestingly, we had expected (based purely on intuition), that Cars would elicit higher rates of descriptive color modification than Toys. However, participants in that condition often described the brand and model of the cars (e.g., ‘the sedan’ or ‘the Porsche’) instead of their color.

## Experiment 2

### Methods

#### Participants

A total of 90 native English speakers over the age of 18 were recruited from Prolific. These participants had not taken part in the previous experiment. All participants reported normal or corrected-to-normal vision and informed consent was obtained prior to testing.

#### Materials and Procedure

We used the experimental design by Rohde and Rubio-Fernandez (2022) to assess listeners’ noun prediction following truncated sentences with color adjectives as the critical condition (e.g., *Click on the green...*) and number as a control (e.g., *Click on the two...*). For each trial (N=42), participants were shown two images of different lexical categories (for sample displays, see Fig. 3) and asked to click on the most likely continuation of the sentence above the pictures.

We compared three pairs of lexical categories: Cars vs Appliances (which do not differ in color typicality but differ in strength of color association), Clothes vs Appliances (which do not differ in color typicality but differ in strength of color association), and Fruits vs Clothes (which differ in color typicality, since fruits often have prototypical colors whereas clothes do not). As such, Fruits vs Clothes (which was the condition originally tested in Rohde and Rubio-Fernandez, 2022) served as our baseline comparison. We did not test the category Toys because it elicited similar rates of color modification as Cars in Experiment 1.

We predicted that participants would select clothes instead of fruits in the baseline condition, replicating the results of Rohde and Rubio-Fernandez (2022). If the original results extend from color typicality to strength of color association, participants should also select the color-associated categories in the two critical pairs (i.e., cars in Cars vs Appliances and clothes in Clothes vs Appliances). Finally, the preference for color-associated categories should be observed in the Color condition but not in the Number condition.

### Results

The data were analyzed with the lme4 package (Bates, Kliegl, Vasishth and Baayen, 2015) from R (R Core Team, 2021). Using logistic mixed effects regression, we modelled the binary outcome variable of Color-Associated Response (1=Category with stronger color association, 0=Category with weaker color association) with Category Pair (Cars vs Appliances, Clothes vs Appliances, Clothes vs Fruits) and Description (Color vs Number) as fixed effects, Fruit vs Clothes as the reference level, and the maximal random effect structure by Participants and Items (Barr et al., 2013).

In line with our predictions, participants selected the category with the stronger color association at comparable rates in the baseline condition (color typicality effect replicating the results of Rohde and Rubio-Fernandez, 2022) than in the two critical conditions (color association effect), with the difference not reaching significance (both  $p$ 's > .05; for the full model output, see Table 2). Also as expected, there was a significant effect of Description ( $p$ =.0006), whereby the lexical category with the stronger color association was selected more often following color descriptions than number descriptions (Fig. 4).



Figure 3: Sample Clothes vs Appliances displays from Experiment 2 from the Color (top) and Number (bottom) conditions.

Table 2: Model output from Experiment 2

Fixed Effect	Coefficient	SE	P-value
Cars vs Appliances	.5998	.3152	.0571
Clothes vs Appliances	.5160	.3051	.0908
Description	-1.011	.2964	.0006
Cars vs Appliances x Description	.1450	.3941	.7129
Cars vs Appliances x Description	-.3557	.3804	.3497

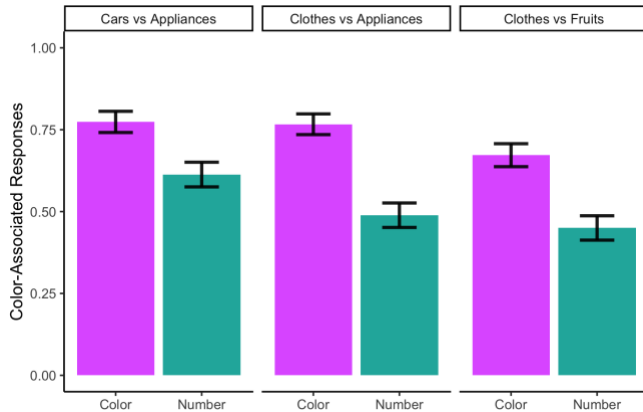


Figure 4: Average rates of Color-Associated Responses (i.e. cars in Cars vs Appliances and clothes in Clothes vs Appliances and Clothes vs Fruits) across the three Category Pairs and two Descriptions in Experiment 2.

## Discussion

Previous referential communication studies have shown color typicality effects on the production and comprehension of color adjectives (Sedivy, 2003, 2004; Westerbeek et al., 2015; Rubio-Fernandez, 2016; Kreiss & Degen, 2020; Long et al., 2021; Rohde & Rubio-Fernandez, 2022). The present study contributes to this growing literature by showing that the relative strength with which a given lexical category is associated with color (e.g., clothes are more strongly associated to color than appliances) influences the production and comprehension of color adjectives when referring to that category, above and beyond color typicality.

Our results show a correspondence between speakers' use of descriptive color modification for different lexical categories (Experiment 1) and listeners' anticipation of upcoming content following color descriptions (Experiment 2). Clothes was the lexical category that elicited the highest rates of descriptive color modification, at more than twice the average rate observed for Toys and Cars (see Fig. 2). This pattern of results is in line with the results of Rubio-Fernandez (2016, 2019), who observed descriptive color modification for clothes but not for geometrical shapes in monochrome displays—where color is a highly inefficient visual cue. However, while considerably lower than Clothes, Toys and Cars still elicited higher rates of descriptive color adjectives than Fruits with predictable colors, whereas Appliances (which are more weakly associated to color) elicited comparably low rates to Fruits.

In conclusion, our results confirm that strength of color association—and not just color typicality—significantly affects the production and comprehension of descriptive color adjectives for different lexical categories. Given the intuitive basis on which we selected the lexical categories tested here, future studies should try to operationalize and further investigate the notion of 'strength of color association.' For example, are lexical categories with more

color variability always more strongly associated to color? Or can color be a central property of a lexical category whose members do not vary greatly in color? For example, cars have greater color variability than appliances, which might explain why they are more often described by color. However, make-up products have less color variability than clothes, but color is a central property of both categories.

Finally, given the observed correspondence between color production and color comprehension, our findings also highlight the need to account for lexical effects when developing pragmatic accounts of efficient referential communication (Rubio-Fernandez, 2019, 2021; Rubio-Fernandez et al., 2021) and computational models linking the production of descriptive color adjectives to the listener's visual search for the referent (Tourtouri et al., 2019; Jara-Ettinger & Rubio-Fernandez, 2021).

## References

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H., (2015). Parsimonious mixed models. *arXiv preprint, arXiv:1506.04967*
- Clarke, A. D., Elsner, M., & Rohde, H. (2013), Where's Wally: The influence of visual salience on referring expression generation. *Frontiers in Psychology*, 4, 329.
- Gatt, A., Krahmer, E., Van Deemter, K., & van Gompel, R.P. (2017), Reference production as search: The impact of domain size on the production of distinguishing descriptions. *Cognitive Science*, 41, 1457–1492.
- Jara-Ettinger, J., & Rubio-Fernandez, P. (2021). The social basis of referential communication: Speakers construct physical reference based on listeners' expected visual search. *Psychological Review*, 129(6), 1394–1413.
- Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, 37, 395-411.
- Kreiss, E., & Degen, J. (2020). Production expectations modulate contrastive inference. *Proceedings of the Annual Meeting of the Cognitive Science Society*.
- Long, M., Moore, I., Mollica, F., & Rubio-Fernandez, P. (2021). Contrast perception as a visual heuristic in the formulation of referential expressions. *Cognition*, 217, 104879.
- Long, M., Rohde, H., & Rubio-Fernandez, P. (2020). The pressure to communicate efficiently continues to shape language use later in life. *Scientific Reports*, 10(1), 8214.
- R Core Team. (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria.
- Rehrig, G., Cullimore, R. A., Henderson, J. M., & Ferreira, F. (2021). When more is more: Redundant modifiers can facilitate visual search. *Cognitive Research: Principles and Implications*, 6, 1-20.

- Rohde, H., & Rubio-Fernandez, P. (2022). Color interpretation is guided by informativity expectations, not by world knowledge about colors. *Journal of Memory and Language*, 127, 104371.
- Rubio-Fernandez, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, 7, 153.
- Rubio-Fernandez, P. (2019). Overinformative speakers are cooperative: Revisiting the Gricean Maxim of Quantity. *Cognitive Science*, 43(11), e12797.
- Rubio-Fernandez, P. (2021). Color discriminability makes over-specification efficient: Theoretical analysis and empirical evidence. *Humanities and Social Sciences Communications*, 8(1), 1–15.
- Rubio-Fernandez, P., Mollica, F., & Jara-Ettinger, J. (2021). Speakers and listeners exploit word order for communicative efficiency: A cross-linguistic investigation. *Journal of Experimental Psychology: General*, 150(3), 583.
- 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.
- Sedivy, J. C. (2004). Evaluating explanations for referential context effects: Evidence for Gricean mechanisms in online language interpretation. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 345–364). Cambridge, MA: MIT Press.
- Tourtour, E. N., Delogu, F., Sikos, L., & Crocker, M. W. (2019). Rational overspecification in visually-situated comprehension and production. *Journal of Cultural Cognitive Science*, 3, 175–202.
- Westerbeek, H., Koolen, R., & Maes, A. (2015). Stored object knowledge and the production of referring expressions: The case of color typicality. *Frontiers in Psychology*, 6, 935.
- Wu, S. A., & Gibson, E. (2021). Word order predicts cross-linguistic differences in the production of redundant color and number modifiers. *Cognitive Science*, 45(1), e12934.