



# Word Order Predicts Cross-Linguistic Differences in the Production of Redundant Color and Number Modifiers

Sarah A. Wu,  Edward Gibson

*Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology*

Received 13 April 2020; received in revised form 17 November 2020; accepted 27 November 2020

---

## Abstract

When asked to identify objects having unique shapes and colors among other objects, English speakers often produce redundant color modifiers (“the red circle”) while Spanish speakers produce them less often (“el círculo (rojo)”). This cross-linguistic difference has been attributed to a difference in word order between the two languages, under the incremental efficiency hypothesis (Rubio-Fernández, Mollica, & Jara-Ettinger, 2020). However, previous studies leave open the possibility that broad language differences between English and Spanish may explain this cross-linguistic difference such that English speakers may generally produce more modifiers than Spanish speakers, including redundant ones, irrespective of word order. Here, we test the incremental efficiency hypothesis in a language production task crossing language (English, Spanish) with modifier type (color, number). Critically, number words occur on the same side of the noun in both English and Spanish. If broad language differences are responsible for the higher rate of color word production in English compared to Spanish, then the same effect should hold for number words. In contrast, the incremental efficiency hypothesis predicts an interaction between language and modifier type, due to different ordering for color words but identical ordering for number words. Our pre-registered analyses offer strong support for the incremental efficiency hypothesis, demonstrating how seemingly small differences in language can cause us to describe the world in surprisingly different ways.

*Keywords:* Language production; Cross-linguistic word order; Incremental efficiency; Incremental planning; Referential communication

---

## 1. Introduction

According to Grice’s Maxim of Quantity, to communicate successfully, people should say enough to uniquely identify a referent, but avoid redundancy by saying no more than

---

Correspondence should be sent to Sarah A. Wu, Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar St., Cambridge, MA. E-mail: sarahawu@alum.mit.edu

is necessary (Grice, 1975; Olson, 1970). Interestingly, however, people often use redundant adjectives in referential communication (Engelhardt, Bailey, & Ferreira, 2006; Koolen, Goudbeek, & Krahmer, 2013; Sedivy, 2005), especially redundant color words (Arts, Maes, Noordman, & Jansen, 2011a, 2011b; Brown-Schmidt & Konopka, 2011; Rubio-Fernández, 2016, 2019; Rubio-Fernández & Jara-Ettinger, 2020; Rubio-Fernández, Mollica, & Jara-Ettinger, 2020). For instance, people will often produce both color and shape words to label objects in Fig. 1, for example, “the red circle,” even though the description “the circle” is sufficient. Even more interestingly, Rubio-Fernández et al. (2020) observed that Spanish speakers tend to produce fewer redundant color words than English speakers. Thus, while an English speaker might refer to the boxed object in Fig. 1 as “the red circle,” a Spanish speaker is more likely not to use a color word and instead call it “el círculo,” not “el círculo rojo.”

To explain this finding, Rubio-Fernández et al. (2020) appealed to the different word orders for adjectives and nouns across the two languages. In English, color adjectives are prenominal and come before their head nouns, whereas in Spanish they are postnominal and come after. They propose the *incremental efficiency hypothesis*, adopting a listener-centric view in which a speaker attempts to produce referential descriptions that optimize the ease of a listener’s visual search for the referential target. According to this hypothesis, speakers of a prenominal language such as English tend to produce redundant color adjectives because color cues may facilitate listeners’ visual search.

However, all previous studies that demonstrate higher rates of redundant color word production in English than in Spanish leave open the possibility that English speakers may simply produce more modifiers in general, including redundant ones. In particular, English speakers may generally be more verbose, and this could lead them to produce more redundant modifiers than Spanish speakers independent of the differing word order of the languages. As there currently exists no corpus of English and Spanish speakers describing the same displays that would provide data about general rates of modifier usage, this possibility remains open. The goal of the current study was to evaluate a task that more stringently tests the incremental efficiency hypothesis by providing a novel control against such broad language or cross-cultural differences.

Here, we examined a case where word order is the same in both languages: the domain of number words. Crucially, number is also used redundantly in English and is not context-sensitive, like color (Brown-Schmidt & Konopka, 2011). Because number words are prenominal in both languages, they provide a control that allows for an incrementality-based interpretation of the difference in production of redundant color words. If factors other than word order caused the difference in color word production between English and Spanish in previous experiments, then the same difference should hold for number words. In contrast, the incremental efficiency hypothesis predicts an interaction between language and modifier type such that English speakers should use more redundant color modifiers than Spanish speakers, but there should be no difference in the usage of redundant number modifiers.

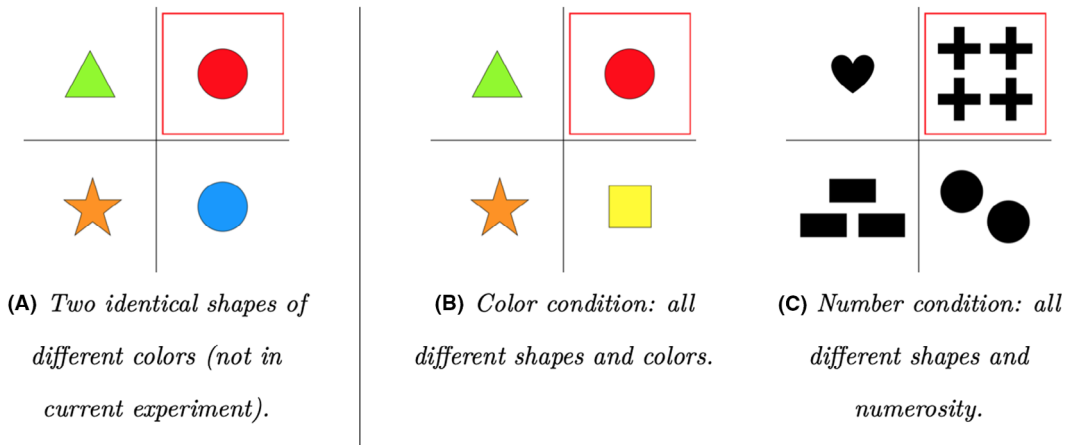


Fig. 1. Sample object displays and experiment materials. In (A), there are four objects including two circles. Suppose that a speaker wants a listener to pick the red circle, and the speaker knows that the listener can see the four objects, but the listener does not have access to their relative positions. The speaker can specify the color (red) along with the shape of the object (circle) to disambiguate the two circles. In (B), all four objects differ in both color and shape. Hence, a listener could pick the correct object if the speaker referred only to the shape of the object (“Pick the circle.”). In (C), all four quadrants consist of groups of different numbers of different shapes.

## 2. Methods

The goals of this preregistered experiment<sup>1</sup> were to (a) replicate previous findings that English speakers use color adjectives significantly more often than Spanish speakers in situations where color words are redundant and (b) test the incremental efficiency hypothesis for redundant number adjectives, which are prenominal in both languages and thus should not exhibit any significant difference in usage. The incremental efficiency hypothesis therefore predicts an interaction between language and modifier type: English speakers should produce more color words than Spanish speakers, but there should be no difference between languages for number words.

### 2.1. Participants

In a pilot version of the experiment, 80 English speakers and 120 Spanish speakers were recruited from Amazon Mechanical Turk and compensated at a rate of \$10/hour. Participants had to have a prior Human Intelligence Task (HIT) approval rate of at least 95%. To obtain US-based Spanish speakers on the platform, which is not yet available in a country where Spanish is dominant, we posted the survey in Spanish and evaluated proficiency in a similar manner to Robenalt and Goldberg (2016). Only participants who self-reported a proficiency of 85% or higher and who actually answered in Spanish were included in the analysis. About two-thirds of pilot participants met these criteria; hence, the 50% greater sample size for Spanish speakers. We found a larger difference in color word usage than in number word usage between the two languages, which shows preliminary support for

the incremental planning hypothesis, but the interaction was not statistically significant. To increase statistical power, we recruited five times as many participants for the preregistered version of the experiment: 400 English speakers and 600 Spanish speakers.

## 2.2. *Design and materials*

The design and materials were adapted from Rubio-Fernández et al. (2020). More specifically, our study was identical to the four-shape condition in Experiment 2 except that it was conducted online instead of being run in-person. Each trial consisted of a  $2 \times 2$  grid of four figures, each of which was one of ten different geometric shapes: circle, cross, diamond, heart, oval, rectangle, square, star, sun, or triangle. Participants were placed in one of two conditions that tested either for color or number modifier usage, and the design was the same in both conditions. In the color condition (e.g. Fig. 1b), the figures were colored with one of ten different colors: black, blue, brown, green, gray, orange, pink, purple, red, or yellow. In the number condition (e.g. Fig. 1c), all figures were colored black and each quadrant contained one, two, three, or four shapes in some random permutation. The target group in these conditions had two, three, or four objects, so it was never the single object. The same stimuli and conditions were used for both English and Spanish-speaking participants.

## 2.3. *Procedure*

In each grid, one of the four quadrants was boxed in red as shown in Figs. 1b and c. Participants were asked to label the figure or group of figures boxed in red such that someone else could later use their response to identify the correct one from the four quadrants. The instructions stated that all the figures in each question had different shapes, and that the person doing the identification task later would not have seen any of the figures beforehand. All participants saw a random set of five trials from a total of 20 generated for each condition. We chose five for the number of trials because participants tend to be very consistent in their use of modifiers across trials, an effect observed in our pilot experiment and also reported in other studies (Rubio-Fernández, 2019; Tarenskeen, Broersma, & Geurts, 2015).

## 2.4. *Data analysis*

Participants' responses on each trial of the experiment were coded based on whether or not they included a redundant color or number modifier in their label. A logistic mixed-effects model was fit to the trial-by-trial modifier use data using the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015). The model included a fixed effect of the interaction between language (contrast-coded: Spanish =  $-0.5$ , English =  $0.5$ ) and modifier type (dummy-coded: number = 1, color = 0), as well as random intercepts for the items and participants with a by-item slope for language. We dummy-coded modifier type so that our first analysis would be an attempted replication of the redundant color usage effect across the two languages.

### 3. Results

After filtering the data for missing and invalid answers, 977 English responses (from 194 participants) and 853 Spanish responses (from 171 participants) were kept for testing color words, and 948 English responses (from 190 participants) and 815 Spanish responses (from 164 participants) were kept for testing number words. The means and 95% confidence intervals for the four groups are illustrated in Fig. 2. English speakers used color adjectives in 81.1% (95% CI: [78.6, 83.5]) of color trials, which was significantly higher than Spanish speakers who used color adjectives in 62.6% (95% CI: [59.4, 65.9]) of trials ( $\beta = 0.625$ ;  $p = .00956$ ). English speakers used number modifiers on 84.2% (95% CI: [81.9, 86.5]) of trials, while Spanish speakers used them on 81.6% (95% CI: [78.9, 84.3]) of trials. Most importantly, the model showed a significant interaction between language and modifier type ( $\beta = -1.111$ ;  $p = 0.0115$ ), indicating support for the incremental efficiency hypothesis. A statistical power analysis conducted with the *simR* package in R (Green & MacLeod, 2016) revealed a power of 84%, showing that our study sample size was sufficient and appropriate for the observed results.

### 4. Discussion

Our study replicated the result that English speakers produce more redundant color modifiers than Spanish speakers. Furthermore, we found no difference in the number of redundant number modifiers that were produced across English and Spanish speakers. This matches the predictions of the incremental efficiency hypothesis (Rubio-Fernández et al., 2020), which states that speakers try to choose referential descriptions that make the referent easy for listeners to find. Accordingly, English speakers often include redundant color words to cue listeners' visual search before they encounter the noun, and both English and Spanish speakers often include redundant number words as similar cues. In

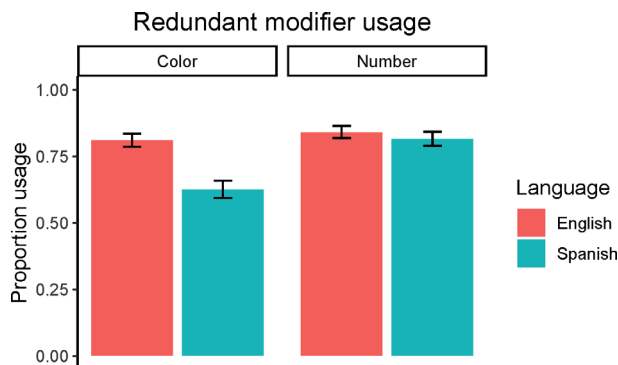


Fig. 2. Proportion usage of redundant color (left) and number (right) modifiers by language, which is indicated by color. Error bars represent 95% bootstrapped confidence intervals.

contrast, Spanish speakers produce fewer redundant color words because such modifiers are produced after the noun and are therefore less useful (efficient) for listeners in finding the target object in the display.

Results from previous studies of redundant modifier production leave open the possibility that broad language or cultural differences between English and Spanish may explain the difference in usage such that English speakers may be more verbose than Spanish speakers irrespective of modifier type. However, such potential factors do not explain the pattern of data here because they incorrectly predict the same effect for both color and number words. Thus, the difference in color word production found in our study and in others (Rubio-Fernández, 2016, 2019; Rubio-Fernández et al., 2020) is likely due solely to the word order difference between English and Spanish.

The incremental efficiency hypothesis is a listener-centric view in the sense that the speaker is motivated by considerations from the listener's perspective. A speaker-centric theory can also explain the observed data. That is, it is also possible that the planning of referential descriptions may be driven by the availability of particular referential components of the display, from the speaker's point of view, rather than in consideration of what might be easiest for the listener (Ferreira & Dell, 2000). Although speaker-based versus listener-based proposals (e.g., Jaeger, 2010) are hard to distinguish in English language production, Zhan and Levy (2018) provide evidence from the production of Mandarin classifiers that speakers produce referential descriptions based on the availability of particular words, and not for the ease of comprehension for listeners. We therefore propose a speaker-centric theory for the production of referring expressions, which we will call the *incremental planning hypothesis*, that can also explain the observed cross-linguistic differences in rates of redundant modifier usage. According to this hypothesis, speakers produce referential descriptions incrementally (Pechmann, 1989) by choosing features that are most salient about the target. To produce an unambiguous description of an object to be identified by the listener, the speaker is sensitive to both the salience of visual features of the object they are referring to and the word order of the language they are speaking (Brown-Schmidt & Konopka, 2008). The speaker plans by attempting to produce words for features that both distinguish the target from competitors and are visually salient, until they are satisfied that the object can be uniquely identified. Thus, in English, people may produce a color word to describe an object among a display of several different colored objects (Belke, 2006), followed by a head noun, the easiest of which is shape, which results in redundancy (e.g., “the red circle” for Fig. 1). In Spanish, in contrast, people produce the head noun first to describe the object (“el círculo”), after which they may notice (perhaps subconsciously) that this description is sufficient to disambiguate the object from its competitors in the display, and so they may finish their description there. Like the incremental efficiency hypothesis, the incremental planning hypothesis predicts that more redundant modifiers will be produced for languages in which the modifiers are prenominal than for languages in which they are postnominal. This is the data observed in our study—that English speakers produce more redundant color modifiers than Spanish speakers because color is prenominal rather than postnominal in English, while rates are similar for number modifiers which have the same ordering

in both languages. Although both hypotheses are plausible, our results do not distinguish the two and future work is needed to tease them apart.

It is interesting to note that the absolute rates of color word usage largely differed between our study and previous studies with the same design. We observed that Spanish speakers produced redundant color modifiers in over 60% of all trials on average, while Rubio-Fernández et al. (2020) reported less than 10%. There are at least three ways in which our studies differed. First, our experiments were conducted with written responses rather than spoken responses. As a result, participants could always edit their initial answers without the experimenter knowing, which may have biased them to subsequently add modifiers. Second, our experiments were conducted through an online platform where participants can be rejected for not doing sufficient quality work. While no rejections were made in our study, many participants typically do hundreds of tasks each week, so they may have been motivated to take more conservative strategies in general such as producing redundant modifiers in our task. Third, our data were collected in the United States, so it is likely that many of our participants were English–Spanish bilinguals. This may have driven up their rates of redundant color word production.

All three of these factors, which stem from the unique features of online crowdsourcing and specifically the Amazon Mechanical Turk platform, may have contributed to the higher absolute rates of redundant modifier usage observed in this study compared to related studies. The difference is especially prominent among Spanish-speaking participants, which may account for the smaller effect size found in the analysis. As a result of this, combined with the fewer number of trials per participant, a larger sample size was required to achieve sufficient statistical power. Despite the unique features of our methodology, the critical between-language comparison still stands and replicates previous findings. The observed rates of number word production confirm that the effect is due to differing word order between English and Spanish, even for bilinguals.

These data thus provide strong evidence for the incremental efficiency hypothesis (as well as the alternative production-based incremental planning hypothesis) over other explanations of redundant color word production. Furthermore, these results demonstrate how seemingly small differences in language that we speak—here, just the order of adjectives relative to nouns—can cause us to describe the world in surprisingly different ways.

## **Acknowledgments**

We thank Mika Braginsky for instruction in data analysis, Rachel Ryskin for assistance in data analysis and valuable feedback on early versions of the manuscript, and Mauricio Garcia for assistance in translations. We also thank Ruud Koolen and two other anonymous reviewers for their helpful comments and suggestions. This research work was supported by the MIT Department of Brain and Cognitive Sciences and NSF Linguistics Program Award 1534318.

## Open Research badges



This article has earned Open Data and Open Materials badges. Data and materials are available at <https://osf.io/mpvah/>.

## Note

1. Preregistered at <https://osf.io/xmuwe>. All materials, data, and statistical analyses available at <https://osf.io/mpvah/>.

## References

- Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011a). Overspecification facilitates object identification. *Journal of Pragmatics*, 43(1), 361–374. <https://doi.org/10.1016/j.pragma.2010.07.013>
- Arts, A., Maes, A., Noordman, L. G. M., & Jansen, C. (2011b). Overspecification in written instruction. *Linguistics*, 49(3), 555–574. <https://doi.org/10.1515/ling.2011.017>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Belke, E. (2006). Visual determinants of preferred adjective order. *Visual Cognition*, 14(3), 261–294. <https://doi.org/10.1080/13506280500260484>
- Brown-Schmidt, S., & Konopka, A. E. (2008). Little houses and casas pequeñas: Message formulation and syntactic form in unscripted speech with speakers of English and Spanish. *Cognition*, 109(2), 274–280. <https://doi.org/10.1016/j.cognition.2008.07.011>
- Brown-Schmidt, S., & Konopka, A. E. (2011). Experimental approaches to referential domains and the on-line processing of referring expressions in unscripted conversation. *Information*, 2(2), 302–326. <https://doi.org/10.3390/info2020302>
- Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, 54(4), 554–573. <https://doi.org/10.1016/j.jml.2005.12.009>
- Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, 40(4), 296–340. <https://doi.org/10.1006/cogp.1999.0730>
- Green, P., & MacLeod, C. J. (2016). simr: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>
- Grice, H. P. (1975). Syntax and Semantics. In P. Cole & J. L. Morgan (Eds.), *Speech Acts* (Vol. 3, pp. 41–58). New York: Academic Press. [https://doi.org/10.1163/9789004368811\\_003](https://doi.org/10.1163/9789004368811_003)
- Jaeger, T. F. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cognitive Psychology*, 61(1), 23–62. <https://doi.org/10.1016/j.cogpsych.2010.02.002>
- Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, 37(2), 395–411. <https://doi.org/10.1111/cogs.12019>
- Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, 77(4), 257–273. <https://doi.org/10.1037/h0029436>
- Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, 27(1), 89–110. <https://doi.org/10.1515/ling.1989.27.1.89>



- Robenalt, C., & Goldberg, A. E. (2016). Nonnative speakers do not take competing alternative expressions into account the way native speakers do. *Language Learning*, 66(1), 60–93. <https://doi.org/10.1111/lang.12149>
- Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, 7, 153. <https://doi.org/10.3389/fpsyg.2016.00153>
- Rubio-Fernández, P. (2019). Overinformative speakers are cooperative: revisiting the Gricean maxim of quantity. *Cognitive Science*, 43(11), e12797. <https://doi.org/10.1111/cogs.12797>
- Rubio-Fernández, P., & Jara-Ettinger, J. (2020). Incrementality and efficiency shape pragmatics across languages. *Proceedings of the National Academy of Sciences*, 117(24), 13399–13404. <https://doi.org/10.1073/pnas.1922067117>
- Rubio-Fernández, P., Mollica, F., & Jara-Ettinger, J. (2020). Speakers and listeners exploit word order for communicative efficiency: A cross-linguistic investigation. *Journal of Experimental Psychology: General*.
- Sedivy, J. C. (2005). 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* (p. 345). Cambridge, MA: MIT Press.
- Tarenskeen, S., Broersma, M., & Geurts, B. (2015). Overspecification of color, pattern, and size: salience, absoluteness, and consistency. *Frontiers in Psychology*, 6, 1703. <https://doi.org/10.3389/fpsyg.2015.01703>
- Zhan, M., & Levy, R. (2018). Comparing Theories of Speaker Choice Using a Model of Classifier Production in Mandarin Chinese. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers) (pp. 1997–2005). New Orleans, Louisiana: Association for Computational Linguistics. <https://doi.org/10.18653/v1/N18-1181>.