Color semantics in language predict color associations in blind and sighted individuals.

Anonymous CogSci submission

Abstract

Colors are strongly associated with certain semantic dimensions (e.g. red is hot, blue is cold). Many of these associations are grounded in our visual perception of the world around us, but blind people can reproduce some of these associations, which suggests color semantics can also be learned from language. How are these color semantics represented in written and spoken language? And how does our use of language align color semantics between individuals? We apply a projection method to word embeddings trained on large corpora of spoken and written text to identify color-semantic associations represented in language. We show that these projections are predictive of color-semantic ratings collected from blind and sighted individuals, but that the effect size varies with embedding training corpus. Finally, we examine how color-semantic associations might be represented in language by training word embeddings on corpora from which various sources of color-semantic information are removed.

Keywords: language; semantic features; distributional semantics

Introduction

Much of what we know about the world we learn by perceiving and interacting with it, and consequently we often talk about knowledge as if it can only be acquired through personal experience. This has sometimes led to the presumption that people who lack certain experiences (e.g. because they are blind or deaf) do not understand the nature of these experiences. It was long thought, for instance, that blind people could not understand the concept of color (both John Locke and David Hume were of this opinion). Evidence from behavioral studies, however, suggests that blind participants can in fact distinguish between cool and warm colors (Shepard & Cooper, 1992) and that some are able to rate color similarities with enough accuracy that multidimensional scaling of the pairwise similarities yields an arrangement that resembles the color wheel (Marmor, 1978; Saysani, Corballis, & Corballis, 2018). Most recently, Saysani, Corballis, and Corballis (2021) collected semantic differential ratings for color words from both blind and sighted individuals and used multidimensional scaling to demonstrate that there is considerable variability between blind individuals and that some, but not all, blind participants generate semantic differentials that are highly similar to the ones generated by sighted people. Since blind participants have no means of directly perceiving color associations, it perhaps seems obvious that they learn color-semantic knowledge from language. *How* color-semantic information is represented in spoken and written language–and to what extent language, rather than visual perception, aligns color semantics between individuals–is, however, not obvious. Are color semantics conveyed explicitly? Are they conveyed through simple cooccurrences? Or are color semantics encoded in more complex semantic structures–a web of associations that we can derive color semantics from?

We conduct four experiments to further explore these questions: In Experiment 1 we reanalyze data collected by Saysani et al. (2021) using word embeddings (a class of distributional semantics model) to get a quantitative measure of the relationship between participants' color-semantic differential ratings and the color semantics represented in language. In Experiment 2 we replicate our findings from Experiment 1 in a sighted sample, but we also explore whether participants perceive their own color semantics as differing from those of others. In Experiment 3 we replicate our findings in yet another, larger, sighted sample, and also explore whether more exposure to certain kinds of language causes participants' color semantics to be more aligned. In Experiment 4, we test several hypothesized origins of color-semantic information in word embeddings by selectively removing them from the corpus the embeddings are trained on.

Experiment 1: Reanalysis of Saysani et al. (2021)

Word Embedding Projections

Using word embeddings, we draw an axis from one end of a semantic dimension (e.g. *hot*) to the other end (e.g. *cold*) and then project the word embedding for each color onto that axis (see Grand, Blank, Pereira, & Fedorenko, 2018 for a discussion of this projection method, but note the method we use here differs slightly in that we normalize the semantic axes before projecting the color embeddings). This provides us with a *relative* measure of word similarity, taken along the semantic dimension's axis, that we can use to predict human ratings of color associations.

Word embeddings are trained on large text corpora, and as such the semantic information they capture tends to reflect the contents of the corpus they are trained on. Using the projection method, we computed color associations in embeddings trained on four different text corpora: The Common Crawl and Wikipedia (Grave, Bojanowski, Gupta, Joulin, & Mikolov, 2018), the OpenSubtitles corpus (Van Paridon &

Thompson, 2020), and the subcorpora of the Corpus of Contemporary American English (COCA; subcorpora are fiction, news, academic texts, spoken texts, and magazine articles).

Method

Participants

Saysani et al. recruited 32 participants, 20 of whom had normal, trichromatic vision. The remaining 12 were congenitally blind, with no residual experience of vision.

Design and Procedure

Participants were asked to rate each of nine color terms (red, orange, yellow, green, blue, brown, purple, black, and white) on 17 semantic dimensions, each defined by two antonyms placed at the poles of a seven-point Likert scale (happy–sad, calm–angry, submissive–aggressive, relaxed–tense, exciting–dull, selfless–jealous, active–passive, like–dislike, alive–dead, fast–slow, new–old, unripe–ripe, soft–hard, light–heavy, fresh–stale, clean–dirty, and cold–hot).

Results

The main finding reported by Saysani et al. (2021) was that multidimensional scaling solutions were more variable between blind participants than between sighted participants. Comparing intraclass correlations for the blind (.35, 95% CI [.29, .42]) and the sighted (.49, 95% CI [.43, .55]) groups also suggests the blind participants are more variable than the sighted participants. Between-participant variability does not mean however that there is no common variance in individual participants' scores that can be predicted from a common measure.

Using a Bayesian linear mixed-effects model with weakly regularizing priors (Yarkoni & Westfall, 2016), we regress word embedding projections onto participants' color-semantic association ratings, while adjusting for dimension word frequency, dimension word concreteness, and how often a given color was provided as response to cueing with a given semantic dimension pole word in the Small World of Words dataset (SWOW; De Deyne, Navarro, Perfors, Brysbaert, & Storms, 2019). The model accounts for random variability by including participant-, color-, and dimension-level random intercepts but also participant-level random slopes for embedding projections, SWOW labels, dimension word frequency and dimension word concreteness, as well as colors and dimensions (see online supplementary materials for more information about model structure).

Both the SWOW differential scores and the word embedding projections have predictive value for the participants' color-semantic ratings. SWOW differential scores have an estimated standardized effect size of -.25 (95% CI [-.29, -.21]); word embedding projections have an estimated standardized effect size of -.58 (95% CI [-.66, -.50]) in sighted participants and -.37 (95% CI [-.48, -.26]) in blind participants when trained on COCA-fiction, but the effect size estimates are smaller for embeddings trained on other corpora, with COCA-spoken performing worst (see Figure 1 for effect size estimates).

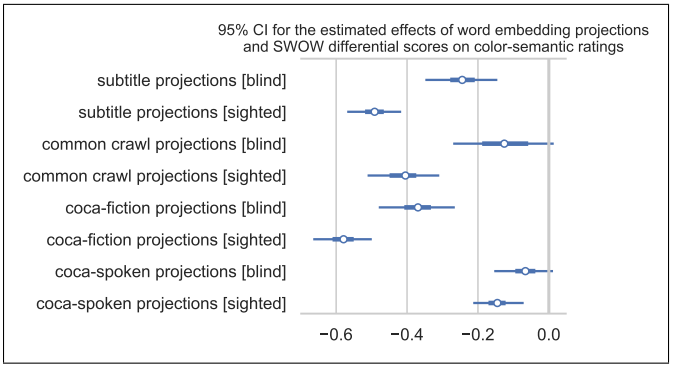

Figure 1: Estimated effects of word embedding projections from various corpora in predicting sighted participants' color-semantic ratings. (Circles are mean estimated effect size, bars represent 95% CIs.)

Discussion

Color-semantic associations in language may arise from our visual perception of the world around us, but they appear to be predictive for both blind and sighted participants' ratings of color semantics. This suggests that even though blind participants are more variable in their color semantics, they learn color semantics from language, aligning their color semantics with those of sighted people.

Differences between word embedding training corpora

Each of the corpora we used yielded color associations that were predictive of the human ratings of color associations, but the corpus that yielded the most predictive associations (i.e. the largest effect size) was the fiction subcorpus of the Corpus of Contemporary American English (COCA-fiction). There are various possible reasons why the fiction subcorpus, specifically, would render a high quality representation of color associations: The fiction subcorpus is relatively high in quality (wide ranging semantic scope, long and coherent sentences, etc.), compared to e.g. the news corpus (which is more limited in semantic scope and contains many short sentences) or the spoken subcorpus (which largely consists of talk radio interviews, etc.). However, what is likely most important is that fiction contains many idiomatic expressions that convey (or are even the primary source of) color associations, such as stating someone is turning blue (when they are cold), green (with jealousy), or red (with anger or embarrassment). These idioms are less common (or absent) in news and academic texts, and may be less consistently (or intelligibly) used in spoken text of the sort that is included in the COCA-spoken subcorpus.

Between-participants variability and self- versus other-ratings

That we are able to use word embeddings to predict some of the variability in participants' color-semantic ratings suggests that they are at least partly informed by a common understanding of color-semantics (a common understanding which is also represented to some extent in the language of large text corpora). Not all variability in participant ratings is predicted by word embeddings however, and the intraclass correlations between participants suggest that there is a considerable amount of variability between participants. Some of this variability may be due to sampling error, but another source of variability could be that participants' color semantics are idiosyncratic, shaped by experiences unique to each individual (e.g. you might think calm is associated with the color yellow because the yoga studio you go to has yellow walls).

Experiment 2: Replication of sighted results

To test whether participants *perceived their own* color-semantic associations as idiosyncratic, we conducted a replication of Experiment 1 with sighted participants, whom we asked to provide not only color-semantic ratings for themselves (*self-ratings*), but also their expectation of color-semantic ratings other participants would provide (*other-ratings*).

Method

Participants

We recruited 30 undergraduate psychology students from the student participant pool at a large research university.

Design and Procedure

Stimuli and procedure were identical to those used in Experiment 1, with two exceptions:

1. The experiment was carried out online rather than in person.
2. We asked participants to provide not only their own color-semantic association ratings, but also the color-semantic associations they expected others to provide.

Results

Using the model structure described in Experiment 1, with the addition of a binary variable describing whether a rating is a self- or an other-rating (and interactions between that variable and the various other variables), we again regress word embedding projections and SWOW differential scores onto participants' color-semantic association ratings.

Self- and other-ratings provided by participants were almost perfectly correlated (Pearson r = .98). Bayesian linear mixed effects modeling showed no difference in self- versus other-ratings (mean estimated effect size .00, 95% CI [-.03, .04]) and embedding projections nor SWOW differential scores were equally predictive of self- and other-ratings (mean estimated interaction for embedding projections and self- vs. other-ratings .01, 95% CI [-.01, .03], mean estimated interaction for SWOW differential scores and self- vs. other-ratings .00, 95% CI [-.02, .02]), for marginal effect sizes in self- and other-ratings see Figure 2.

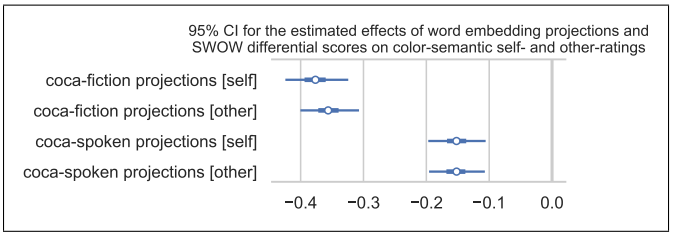

Figure 2: Estimated effects of word embedding projections and SWOW differential scores in predicting sighted participants' color-semantic self-ratings and other-ratings. There is no meaningful difference between self- and other-ratings. (Circles are mean estimated effect size, bars represent 95% CIs.)

Discussion

While some color-semantic associations *may* be idiosyncratic, participants certainly do not perceive their own associations as differing from others'.

Experiment 3: Associating reading measures with fiction-like color semantics

Method

Participants

We recruited 100 undergraduate psychology students from the student participant pool at a large research university.

Design and Procedure

Stimuli and procedure were identical to those used in Experiment 1, with two exceptions:

1. The experiment was carried out online rather than in person.
2. We asked participants how many hours per week they spend reading fiction and nonfiction text, and had them complete the Author Recognition Test (ART; Stanovich & West, 1989). The ART is meant to assess a participant's exposure to the names of prominent authors and is predictive various other reading-related measures.

Results

We analyzed this experiment using the model described in Experiment 1, with added predictors for the reading measures and their interaction with the word embedding projections from the COCA-fiction corpus. None of the reading measures interact with the word embeddings projections in predicting color-semantic ratings (see Figure 3 for estimated effect sizes).

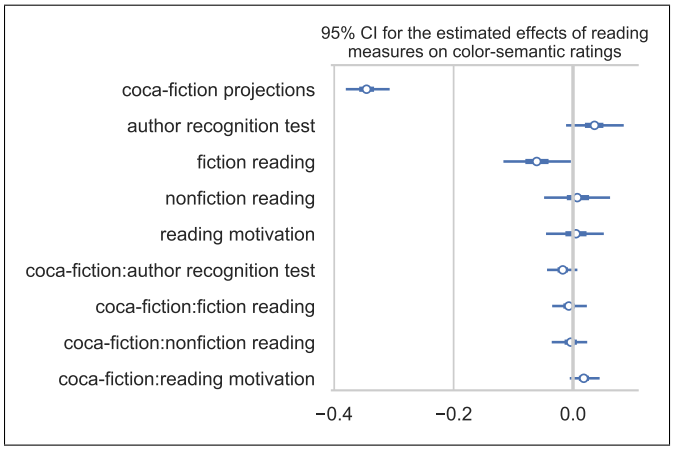

Figure 3: Estimated effects of word embedding projections, various reading measures, and the interactions between projections and reading measures in predicting sighted participants' color-semantic ratings. The predictiveness of fiction-derived word embedding projections does not appear to be higher for participants with more reading exposure. (Circles are mean estimated effect size, bars represent 95% CIs.)

Discussion

Given the predictive power of the COCA-fiction word embedding projections (relative to those based on spoken text or Wikipedia/Common Crawl), it is somewhat surprising that reading more fiction does not cause participants to be more aligned with the word embedding projections. However, it is possible that since our participants are all undergraduate students at a large research university, their shared linguistic and cultural background (reading the same books in school, watching the same tv shows growing up, using the same social media) could mean they are already strongly semantically aligned, essentially creating a ceiling effect that obscures any alignment due to additional exposure to written fiction.

Meta-analytic effect size of embedding projections in predicting color semantics

When we pool data from all three experiments, we can obtain a better estimate of the effect size of embedding projections in predicting the color-semantic ratings of sighted participants. [Insert meta-analytic effect size and discussion?]

Where do the embeddings "learn" their color semantics?

One potential way of finding the source of these color associations (in the word embeddings, at least) is to try to modify the training corpus in such a way that the associations disappear from the embeddings (and the projected associations are no longer predictive of human ratings). If we can identify the sentences in the training corpus that give rise to the color associations, can these sentences tell us whether (and how) humans learn color associations from language?

Experiment 4: Identifying sources of color-semantic information in embedding training corpora

Identifying the sentences that are most informative for color associations in a training corpus is not a trivial problem: there is a computational cost to training word embeddings, so performing an exhaustive search is not feasible. Nevertheless, we can start by testing several "naive" hypotheses:

(a) Color associations are driven by *first-order* cooccurrences, the occurrence of a color word and a semantic dimension word in the same sentence (e.g. "the fire was *red hot*"; color associations in these sentences *can* be explicit, but often are not).

(b) Color associations are driven by *second-order* cooccurrences, the occurrence of color words and semantic dimension words in similar contexts (i.e. color words and semantic dimension words may not cooccur, but the share words that they cooccur with, e.g. "Southern cooking uses *green* tomatoes" and "Southern cooking uses *unripe* tomatoes").

(c) Color associations are driven by cooccurrences between color words and words in the same semantic neighborhood as semantic dimension words (e.g. "The forest was *white* with *snow*", snow being in the same semantic neighborhood as *cold*).

(d) Less direct semantic associations that happen to be very salient. We do not expect these to be represented in the training data in a specific form, but their salience means we should be able to easily elicit them from human participants.

These sources of color-semantic information need not be mutually exclusive; words captured by (c) and (d) may overlap, and all of these words may be part of the set of word described by (b).

Methods

Participants

We recruited 100 undergraduate psychology students from the student participant pool a large research university to provide labels associated with color-semantic dimensions.

Design and Procedure

For the first-order cooccurrence hypothesis, we removed from the COCA-fiction corpus any sentence containing both a color word and one of our semantic dimension words. For the semantic neighborhood hypothesis, we removed from the COCA-fiction corpus any sentence containing one of the 10 nearest neighbors of each semantic dimension word. The second-order cooccurrence hypothesis proved to be difficult to test: The number of sentences containing shared words is vastly larger than the number of sentences containing first-order cooccurrences, so indiscriminately removing all of these shared words (and the sentences they occur in) reduces the size of the training corpus by an order of magnitude, making it infeasible to test this hypothesis without narrowing

down which shared words are most informative for the color-semantic associations.but this reduced the size of the corpus by an order of magnitude, making it impossible to properly contrast embeddings trained on this filtered corpus with the original COCA-fiction embeddings.

For the indirect links through salient words hypothesis, we collected labels for color-semantic associations. Participants were presented with a prompt asking them to provide a label for the pairing of a color and one of the semantic dimension words used in Experiments 1-3 (e.g. "What word comes to mind for *white* and *cold*?"), this was repeated for each of three colors and 34 dimension words, for a total of 102 trials per participant (color groupings were counterbalanced across participants so that all nine colors were presented an equal number of times). We computed name agreement and Simpson's diversity (Simpson, 1949) for the labels participants provided for each color/semantic dimension word pair. We then removed from the COCA-fiction corpus any sentence containing a label provided by at least two participants before training word embeddings on the corpus.

Results

Removing first-order cooccurrences did not meaningfully reduce the effect size of the COCA-fiction word embedding predictions. Removing nearest neighbors and especially removing participant-generated labels for color-semantic associations had a measurable impact however (see Figure 4 for estimated effect sizes).

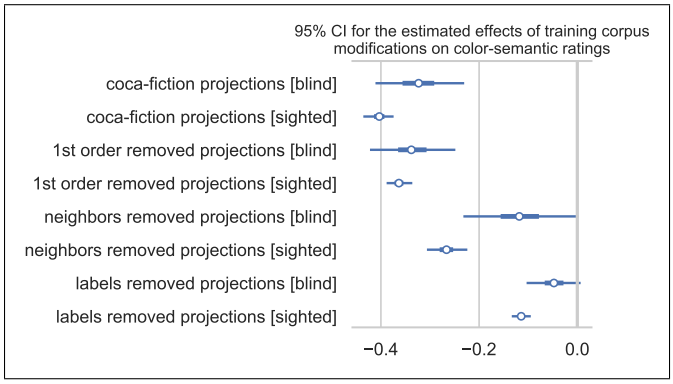

Figure 4: Estimated effects of word embedding projections in predicting blind and sighted participants' color-semantic ratings. (Circles are mean estimated effect size, bars represent 95% CIs.)

Discussion

That removing first-order cooccurrences had no measurable effect is perhaps not surprising: The objective use to train word embedding models is to predict the context a word occurs in. Strict first-order cooccurrence therefore does nothing to drive embedding similarity. That removing participant-generated labels for color-semantic associations is so effective in removing color-semantic information from the embedding projections is striking, since the number of labels generated by at least two participants (the threshold for inclusion in our corpus filtering procedure) was only about 300 words. Future work will explore whether these labels also satisfy the second-order cooccurrence hypothesis, or if the semantic structures that underpin color-semantic associations arise from still higher-order cooccurrence patterns.

Acknowledgments

We would like to thank Armin Saysani and Michael and Paul Corballis for making available the raw data we reanalyzed in Experiment 1.

References

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "small world of words" english word association norms for over 12,000 cue words. *Behavior research methods*, *51*(3), 987–1006.

Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2018). Semantic projection: recovering human knowledge of multiple, distinct object features from word embeddings. *arXiv preprint arXiv:1802.01241*.

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.

Marmor, G. S. (1978). Age at onset of blindness and the development of the semantics of color names. *Journal of experimental child psychology*, *25*(2), 267–278.

Saysani, A., Corballis, M. C., & Corballis, P. M. (2018). Colour envisioned: Concepts of colour in the blind and sighted. *Visual Cognition*, *26*(5), 382–392.

Saysani, A., Corballis, M. C., & Corballis, P. M. (2021). Seeing colour through language: Colour knowledge in the blind and sighted. *Visual Cognition*, 1–9.

Shepard, R. N., & Cooper, L. A. (1992). Representation of colors in the blind, color-blind, and normally sighted. *Psychological Science*, *3*(2), 97–104.

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*(4148), 688–688.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, 402–433.

Van Paridon, J., & Thompson, B. (2020). subs2vec: Word embeddings from subtitles in 55 languages. *Behavior Research Methods*, 1–27.

Yarkoni, T., & Westfall, J. (2016). Bambi: A simple interface for fitting bayesian mixed effects models. *OSF Preprints*.