COG260 Final Report

Kiarash Kianidehkordi
University of Toronto

Kiera Flanagan
University of Toronto

Abstract

We tested the hypothesis that early-stage color terms, as defined by Berlin and Kay's (1969) evolutionary sequence, exhibit lower cross-linguistic variability than later-stage terms. Using focal color data from 110 unwritten languages in the World Color Survey, we computed centroids in CIELAB space for each basic color term and quantified dispersion as the mean ΔE distance of individual focal points from these centroids. Clustering algorithms (k-means and Gaussian Mixture Models) identified perceptual clusters, which were mapped to universal color categories using English-derived and Munsell-based anchors. Across all methods and anchor sets, early-stage terms, such as black, white, and red, showed tight clustering, while later-stage terms, such as purple, pink, orange, and grey, were substantially more dispersed. Permutation tests confirmed that these differences were statistically significant. Thus, results support our hypothesis that early-stage color terms are cross-linguistically more stable, whereas later-stage color terms are more dispersed.

Keywords: Basic color terms, Evolutionary Sequence,

Introduction

Berlin and Kay's foundational study demonstrated that languages show strict limitations on which basic color categories they encode when they have fewer than eleven color terms (Berlin & Kay, 1969). Their work revealed that all languages include white and black; three-term languages include red; four-term languages include either green or yellow; five-term languages include both green and yellow; six-term languages include blue; seven-term languages include brown; and languages with eight or more terms include brown plus some combination of purple, pink, orange, and grey. Out of the 2,048 theoretical combinations of basic color terms, only 22 actually occur, forming a robust ordering: (black, white) < red < (green, yellow) < blue < brown < (purple, pink, orange, grey).

Berlin and Kay interpreted this sequence as an evolutionary model of colour-term development. Earlier stages contain categories that appear in nearly all languages and are therefore treated as historically older, while later stages include categories that are less common and are presumed to be younger. They supplemented the cross-linguistic distribution with internal linguistic evidence to estimate the relative ages of colour terms. Words that were borrowed from other languages or built from multiple morphological parts were treated as newer additions to the lexicon, while colour terms that consisted of unanalyzable native roots were treated as older forms (Berlin & Kay, 1969). This evidence supports the view that late-stage colour terms are historically younger, but it remains indirect and qualitative. Taken by itself, these findings do not determine whether the proposed stages reflect differences in perceptual–cognitive structure or whether they arise from historical, cultural, or communicative pressures.

The World Color Survey (WCS) was undertaken to validate, modify, or challenge these findings using color-naming data from 110 unwritten languages and a standardized array of 329 Munsell chips (Cook et al., 2005). After identifying each language's basic color terms, speakers marked all chips corresponding to each term and then selected the best example, or focal point. Although the naming task resulted in variable category boundaries across languages, the focus task offers an opportunity to examine cognitive predictions linked to the evolutionary sequence.

Previous work has also suggested that the emergence of basic color terms reflects an increasingly fine-grained and optimal partitioning of perceptual color space (Regier et al., 2007). Early-emerging terms are thought to carve up the perceptual space along the most salient and universally shared distinctions, whereas later-emerging terms subdivide already-partitioned regions. Because of this, we expect that color terms that emerge later in the evolutionary sequence will show greater cross-linguistic variability in their focal choices.

Our project tests the hypothesis that colour categories of unwritten languages from earlier stages of the Berlin and Kay (1969) sequence will exhibit significantly lower cross-linguistic variability in focal points than categories from later stages. This expectation reflects the idea that early-stage terms, such as black, white, and red, show greater perceptual–cognitive entrenchment, while late-stage terms, such as purple, pink, orange, and grey, show weaker cross-linguistic stability. An alternative perspective is that evolutionary stage and focal dispersion may not be systematically related. Although our hypothesis builds directly on the evolutionary sequence proposed by Berlin and Kay (1969), there are documented exceptions to this ordering across languages. If basic color terms are indeed universal in some form, these exceptions suggest that the Berlin and Kay sequence may not fully capture the true historical or cognitive development of color lexicons. Under this view, the relationship between focal-point variability and evolutionary stage might not follow the predicted pattern, and a different evolutionary sequence of basic color terms could better account for cross-linguistic variation.

Methods

We analyzed naming and focal color data from the World Color Survey (WCS), which documents color terminology across 110 unwritten languages using a standardized palette of 329 Munsell chips (Cook, Kay, Regier, & Kay, 2005). The code first loads all focal responses using custom helper functions and filters terms by a minimum speaker consensus of ten responses.Each focal chip's Munsell specification was converted to CIELAB coordinates using a Python Munsell-to-CIELAB conversion package, allowing all subsequent computations to take place in a perceptually uniform color space.

To interpret these data relative to universal color categories, we derived a set of perceptual anchor points from two sources. First, the English anchors were taken from Sturges and Whitfield (1995), who experimentally located English basic color foci in Munsell space; these were converted to CIELAB using the same Munsell-conversion package for methodological consistency. Second, we constructed a parallel Munsell-based anchor set whose coordinates were likewise converted into CIELAB. Each anchor was then assigned to an evolutionary stage following Berlin and Kay (1969).

To find the basic color terms from WCS data, we applied two clustering algorithms—k-means and Gaussian Mixture Models (GMMs)—each configured to identify 11 clusters corresponding to the 11 universal categories. For each clustering solution, cluster centroids were labeled by identifying the nearest anchor in CIELAB space, producing a cluster-to-category mapping used to assign every observed focal centroid to a universal color category and evolutionary stage.

For each basic color term in each language, we compiled all focal points, computed their centroid in CIELAB space, and quantified dispersion as the mean ΔE distance of focal responses from that centroid. We then tested our hypothesis—that later-stage categories exhibit greater cross-linguistic variation—by computing the correlation between evolutionary stage and dispersion using Spearman's ρ and estimating significance with a 5,000-iteration permutation test.

The code automatically generates all figures reported in the paper, including (a) violin plots showing the distribution of dispersion values across evolutionary stages and (b) bar plots showing category-level mean variability color-coded by anchor hue. All data processing steps, analyses, and visualizations are fully reproducible with the code included in the Appendix. To refer to visualizations, see Figures 1–7.

Results

Across all analyses, we observed a robust and consistent pattern in cross-linguistic color term variability: later-evolving color categories are substantially more dispersed than early-evolving ones. Early-stage categories, such as black, white, and red—those that appear first in Berlin and Kay's (1969) proposed evolutionary sequence—exhibited strong cross-linguistic agreement, with relatively low mean dispersion values, indicating that speakers of different languages converge closely on these colors. In contrast, later-acquired categories, such as purple, pink, orange, and grey, displayed markedly higher variability, with focal points spread over a larger portion of perceptual color space. This pattern was evident across both clustering methods and anchor sets.

Using k-means clustering with English anchors, early-stage terms had a mean dispersion of 12.21 ΔE units, whereas late-stage terms were almost twice as dispersed at 21.73 ΔE. A parallel pattern emerged with Munsell anchors, where early-stage terms averaged 13.88 ΔE compared to 24.52 ΔE for late-stage terms.

Gaussian Mixture Models produced comparable results: with English anchors, early-stage terms averaged 11.47 ΔE, while late-stage terms reached 21.73 ΔE; with Munsell anchors, early-stage dispersion was 13.88 ΔE versus 24.52 ΔE for late-stage terms.

In all four analytic conditions, the hypothesis that late-evolving color terms are more dispersed was strongly supported. Permutation tests confirmed that these differences were statistically significant, demonstrating that the observed increase in variability for late-stage categories is highly unlikely to have arisen by chance. Together, these results provide compelling evidence for a systematic relationship between evolutionary stage and cross-linguistic dispersion, highlighting the relative universality of early color terms and the greater flexibility and variability of later-emerging categories.

Summary statistics for within-cluster mean dispersion and permutation test significance are reported in Tables 1 and 2, respectively.

Discussions

Our analyses of the World Color Survey data support our original hypothesis, highlighting how early-evolving color categories exhibit tighter cross-linguistic clustering, whereas later-evolving categories are markedly more dispersed.

Basic color terms from unwritten languages mapped onto English and Munsell terms of black, white, and red—those that emerge earliest in Berlin and Kay's (1969) proposed evolutionary sequence—show strong convergence across languages, suggesting that these categories are more perceptually stable.

In contrast, basic color terms mapped onto later-stage categories, such as purple, pink, orange, and grey, display substantially greater variability, indicating that these color terms are perhaps less perceptually constrained and more susceptible to linguistic and cultural influences.

While this pattern was consistent across clustering methods (k-means and Gaussian Mixture Models) and anchor sets (English-derived and Munsell-based), our plots indicate that the resulting clusters found by Gaussian Mixture Models require more nuance. For example, in the GMM category mean dispersion plots (Figures 3 and 7), no cluster was assigned to the grey category; hence, the interpreted clusters are not fully robust across implemented models. Although this outcome is expected, given that the distribution of color terms may be skewed. This may lead to some late-stage categories occasionally disappearing because Gaussian components are allocated according to density in the color space. Dense regions corresponding to early-stage colors, such as red or blue, may capture multiple components, leaving none for less densely populated categories, such as pink or grey.

Both solutions support the conclusion that early color categories are more entrenched across languages, while later categories remain flexible and heterogeneous. However , the observed differences in mean dispersion across anchor sets are attributable to the way clusters are mapped onto universal categories. Clustering algorithms operate purely in perceptual color space and do not inherently “discover” semantic categories; category labels are assigned post hoc based on proximity to predefined anchors. Thus, small changes in anchor coordinates can alter which clusters are designated as early- versus late-stage categories, leading to modest differences in mean dispersion values.

Moreover, several limitations should be considered. Focal-point data can be noisy, and individual variation may inflate dispersion, particularly for late-stage categories that are less frequently attested. Moreso, the assignment of clusters to evolutionary stages relies on English or Munsell anchor sets, introducing potential bias. Alternative anchor systems or different cultural reference points could yield different mappings and stage assignments. Additionally, the WCS dataset primarily samples small-scale, unwritten languages, and late-stage colors tend to have smaller sample sizes, which can amplify apparent variability. Finally, the Berlin and Kay (1969) evolutionary sequence itself may not reflect universal cognitive constraints, but rather cultural, ecological, or communicative pressures that shape color term development.

Taken together, our findings indicate somewhat robust cross-linguistic evidence for our hypothesis : early color categories are tightly clustered, perceptually salient, and cognitively entrenched, whereas later-evolving categories exhibit greater variability and diffuse representation across languages.

Future research can build on these results by examining how later color categories stabilize over time, exploring the impact of language contact, and investigating potential cultural, ecological, or communicative constraints that shape the evolutionary trajectory of color lexicons.

Tables

Table 1: within cluster mean dispersion across evolutionary stage

Method	Early	Late
Kmean(eng)	13.882	24.520
GMM(eng)	11.472	21.735
Kmean(mun)	13.881	24.520
GMM(mun)	12.211	21.734

Table 2: Permutation Test p-values (5,000 Iterations) by Clustering Method and Anchor Set against null distribution

Method	anchors	P-perm
Kmeans	English	0.0
GMM	English	0.0
Kmeans	Munsell	0.0
GMM	Munsell	0.0

Figures

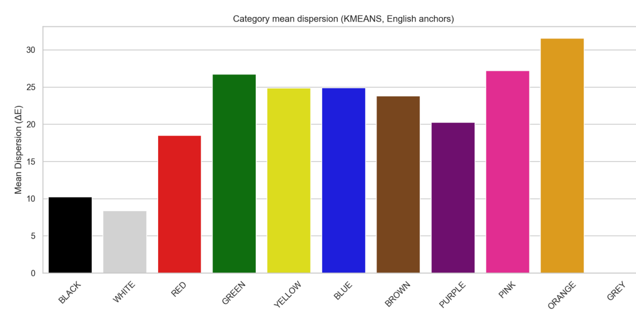

Figure 1: Category mean dispersion for KMeans derived clusters with English anchors.

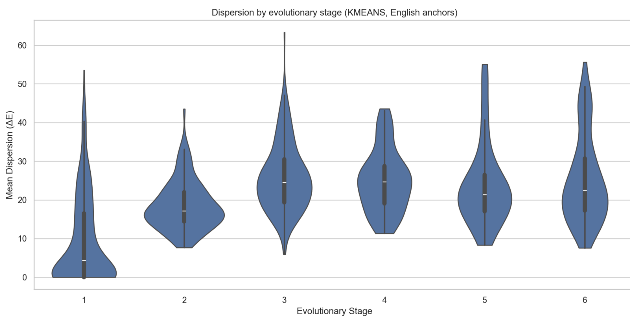

Figure 2: Mean dispersion by evolutionary stage for KMeans derived clusters with English Anchors

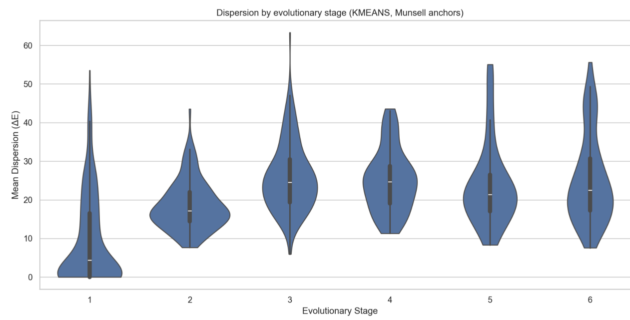

Figure 6: Mean dispersion by evolutionary stage for KMeans derived clusters with Munsell Anchors

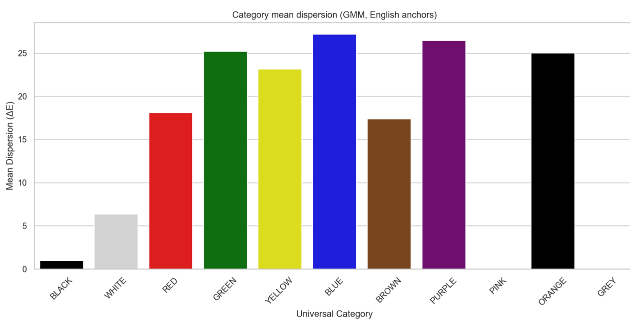

Figure 3: Category mean dispersion for GMM derived cluster with English Anchors

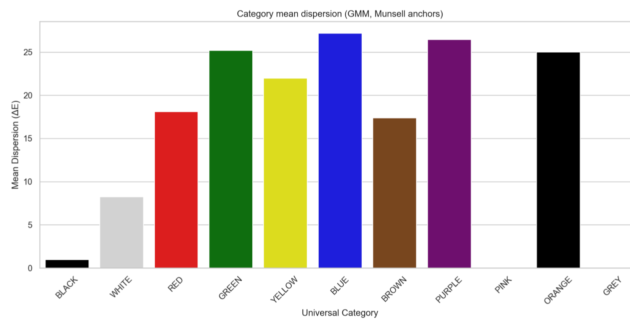

Figure 7: Category mean dispersion for GMM derived cluster with Musell Anchors

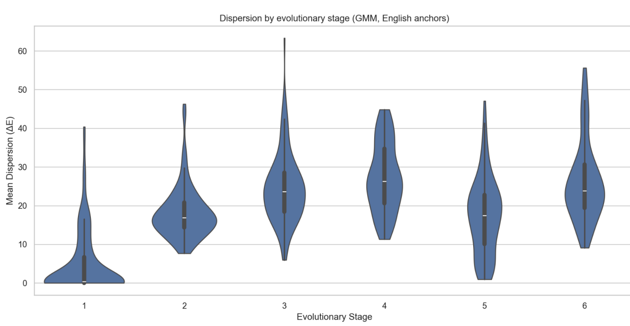

Figure 4: Mean dispersion by evolutionary stage for GMM derived clusters with English Anchors

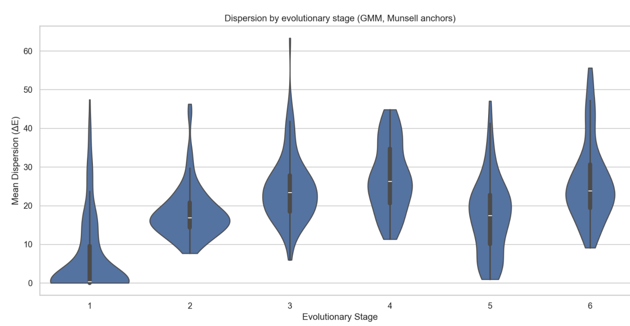

Figure 8: Mean dispersion by evolutionary stage for GMM derived clusters with Munsell Anchors

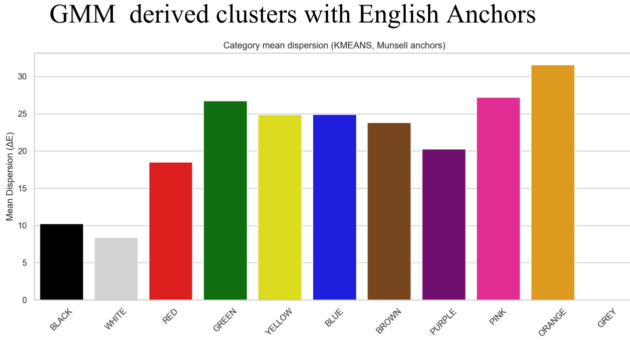

Figure 5: Category mean dispersion for KMeans derived cluster with Munsell anchors

Division of Labour

We divided the work for this report as evenly as possible. Kiera wrote the Introduction. Both authors contributed to the Methods section and the Discussion. Kiarash wrote the Results. For the data analysis step , we used a shared Google Colab environment and wrote the main part of the code in pair-programming sessions.

References

ASTM International. (2018). *Standard practice for specifying color by the Munsell system* (ASTM D1535–18).

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. University of California Press.

Cook, R. S., Kay, P., & Regier, T. (2005). The World Color Survey database: History and use. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in the cognitive sciences*. Elsevier.

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, 104(4), 1436–1441. https://doi.org/10.1073/pnas.0610341104

Sturges, J., & Whitfield, T. W. A. (1995). Locating basic colours in the Munsell space. *Color Research & Application*, 20(6), 364–376. https://doi.org/10.1002/col.5080200605

Appendix

Our analysis script is available online:

https://github.com/kiarashkianid/World-Color-Survey-Evolutionary-sequence-of-basic-color-terms-Data-analysis