COLOR SALIENCE, COLOR TERM EVOLUTION, AND THE ISSUE OF RELATEDNESS

CALEB EVERETT [*1], COLLEEN O'BRIEN [2]

[*] Corresponding Author: caleb@miami.edu
[1]Anthropology, University of Miami, Coral Gables, USA
[2] JCRS, Friedrich Schiller Universität, Jena

Gibson et al. (2017, GEA henceforth) calculated the average surprisal or "entropy" of languages' basic color terms, based on the data of the World Color Survey (Kay et al. 2009). Their innovative work suggests, *inter alia*, that industrialization may yield less entropy and greater efficiency of basic color terms. They obtained the average "surprisal" of each color chip named in the survey, for each language. In their words, "The surprisal score for each color *c* is computed by summing together a score for each word *w* that might have been used to label *c*, which is calculated by multiplying P(w|c) by -log(P(c|w)), the listener's surprisal that *w* would label *c*. We estimate P(c|w) via Bayes Theorem assuming a uniform prior on P(c)." Formulaically:

$$S(c) = \sum_{w} P(w|c) \log \frac{1}{P(c|w)}$$

Via this approach, GEA find that "warmer" Munsell colors (e.g. red and yellow) are encoded with greater efficiency or less surprisal, when contrasted with "cool" colors (e.g. green and brown). This central finding holds across the 110 languages in the WCS plus 3 others tested. The approach can also be used to generate an average "entropy" score for each of these 113 languages by averaging all the color chips' surprisal values in a given language.

We used GEA's code to ascertain languages' average entropy scores, to more carefully examine a secondary claim made in their work: "industrialization… increases color usefulness" and helps to reduce the entropy of color terms in a language. (10785) Color terms may evolve to more efficiently encode the hues associated with foregrounded objects in industrialized cultures, objects which tend to be warm-colored according to GEA's ancillary analysis of thousands of images. GEA note that English and Spanish have the most efficient color term sets in their data, consistent with the industrialization hypothesis. Yet our reanalysis of the WCS data suggests no clear association between industrialization and color terms, given the confound of language relatedness. The entropy scores (average per-color surprisal) of languages range from 3.08 (English) to 5.88 (Eastern Cree) (mean=4.7, s.d.=0.37, n=113). The most "efficient" languages consist primarily of two clusters that deviate from typical efficiency values: a) Indo-European languages and European-based creoles and b) Mesoamerican

languages. At the other end of the efficiency/entropy spectrum, Algonquian languages tend to have high entropy color terms. See Figure 1 for a visual summary of the results across all families. In short, a few language families and one geographic region are over-represented at the two ends of the color-entropy continuum. Indo-European languages could very well have low entropy colors because of industrialization, but we conclude that any claims regarding the causal role of industrialization are difficult to support with the WCS data. The confound of language relatedness, not to mention language contact, cannot be adequately controlled for with WCS-based efficiency data.

As the exception that proves the rule, we describe the case of the Colombian isolate Kamsá. Kamsá has a surprisingly efficient color term system, superficially violating the expectations of GEA's secondary hypothesis. Based on our (CO) own research with Kamsá, however, we discuss how the WCS data were gathered in part or entirely with bilingual Kamsá speakers, given the high rates of Spanish bilingualism since the 1970's. (With bilingualism in Inga prevalent prior to Spanish bilingualism.) We suggest that the efficiency of Kamsá responses may be the result of that Spanish bilingualism, further pointing towards relatedness as a confound in the color-term efficiency findings. It seems quite plausible that industrialization has impacted color terms in the ways GEA suggest. Yet their secondary hypothesis cannot be adequately supported by the WCS data, given the potential role of common ancestry in shaping color term entropy.

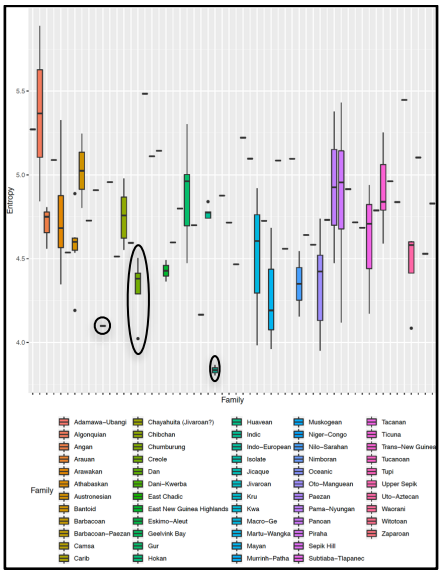

Figure 1. Box plots of color-term entropy scores in the WCS, by language family. Kamsá, creoles, and IE languages (in that order, from left to right) are highlighted.

References

Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S. & Conway, B. (2017). Color naming across languages reflects color use. *PNAS 114*, 10785-10790.

Kay, P., Berlin, B., Maffi, L., Merrifield, W. & Cook, R. (2009) *The World Color Survey.* Stanford: CSLI.