Basic Colour Terms are indispensable

Dimitris Mylonas[1,2], Andrew Stockman[2] and Lewis D. Griffin[1]
[1]Dept. of Computer Science, University College London
[2]Inst. of Ophthalmology, University College London

Colour categorization illuminates the relationship between perception and language. Berlin & Kay[1] identified a set of universal colour categories, the Basic Colours Terms (BCTs) black, grey, white, red, orange, yellow, green, blue, purple, pink and brown. Their criterion was based on multiple factors (e.g. single word terms that are not the name of an object) judged by experts. Others have segregated colour categories on objective behavioural criteria such as frequency and consistency[2,3,4]. Recently colour categories have been studied by measuring their communication effectiveness using information theoretic analysis. For example, 'Alice' is presented with a colour, which she names. 'Bob' then attempts to guess the colour from the name. How many tries will Bob on average take to guess the colour? Assuming Alice names colours like the population average, and Bob guesses optimally, then performance is computable from colour naming data as

$$surprisal(c) := -\sum_{n} P(n|c)\,\log_2 P(c|n)$$

where $P(n\,|\,c)$ is the conditional probability that name *n* will be chosen for a colour *c*, and $P(c\,|\,n)$ that colour *c* was the cause of a naming *n*. Surprisal tends to be higher for cooler than warmer colours[5] reflecting the larger perceptual extents of categories in this region of colour space.

Here, we unite these two approaches with a novel information theoretic measure, which we call *dispensability*, for colour categories that we suspected would predict Basicness. Our measure is an analogue of the surprisal measure. Alice is given a colour name and points at a colour which could give rise to the name. Bob then attempts to guess the name from the indicated colour. How many tries will Bob on average take to guess the name? If Alice chooses colours for names according to colour naming data, and Bob is optimal in his guessing, then performance can be computed as

$$dispensability(n) := -\sum_{c} P(c|n)\,\log_2 P(n|c)$$

We computed dispensability using two colour naming datasets. One was collected using an *online* experiment where 500 English speakers each named freely 20 colours randomly chosen from a range of 600. Eliminating unique names left 7405 responses using 478 names. The other was collected using laboratory presentation of the same stimuli on a calibrated CRT, with each of 10 English speakers freely naming all 600 colours. Eliminating unique names left 4684 responses using 246 names.

Dispensability varies with category and produces a graded scale of Basicness. For example, for yellow the score was 1.39 for the online and 1.28 for the lab datasets; while for mustard the scores were 2.32 and 2.11. Most remarkably, *for both datasets all BCTs had lower dispensability scores than all non-BCTs*. The range of dispensability for basic terms was

1.39-1.90 for the online and 0.81-1.40 for the lab, while for non-basics the ranges were 2.04-4.49 and 1.75-3.25. The lowest dispensability non-BCTs were turquoise and lilac for the online and lime green and beige for the lab datasets.

Dispensability does not identify the BCTs because they are commonly used (e.g. dark green is more frequent in our datasets than white). Rather it works because for each there are colours for which all other names are rarely used – hence the title we have given our measure.

1. Berlin, B., & Kay, P. (1969/1999). *Basic color terms : their universality and evolution*. Stanford, Calif.: Center for the Study of Language and Information.
2. Boynton, R. M., & Olson, C. X. (1987). Locating basic colors in the OSA space. *Color Research & Application*, *12*(2), 94–105.
3. Lindsey, D. T., & Brown, A. M. (2014). The color lexicon of American English. *Journal of Vision*, *14*(2), 17.
4. Mylonas, D., & MacDonald, L. (2016). Augmenting basic colour terms in English. *Color Research & Application*, *41*(1), 32–42.
5. Gibson, E., et al. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, 201619666.