Color Naming in Two Languages

Data Visualization Course Spring`16
Kyle Thayer, Younghoon Kim, and Gabriella Gorsky

Motive

- Extend previous work on mapping different patterns of color naming across languages
  - Previous data sparse, some languages over-represented
  - Participants not all native speakers of language used
- New study for a more fine-grained exploration
  - Gathering denser data from few languages
  - Collecting more demographic information
- New data expands the potential to apply statistical tests
  - Where are the boundaries between color groups?
  - Are there statistically significant differences in color naming trends across languages?
- Prior research shows that language has dynamic effects on color perception
- Other work has explored the interaction between visual--perceptual differences related to language and color naming
- The goal of this study was to continue the exploration of this interaction while adding considerations such as situational lighting conditions and discipline-related color immersion

Study Design

- We created an online “Color Perception” test and administered it on the LabInTheWild platform (labinthewild.org)
- Data Collected:
  - Demographic Information
    - including native language, education level, and situational lighting conditions
  - Color names of 36 colors [see Figure 1]
    - User viewed and named tiles (12 colors at a time)
    - Names were requested in users’ native language
    - Colors tiles chosen from max hue and value edges of RGB color cube
    - One color randomly chosen from 36 segments (bins)
  - Color Sorting [see Figure 2]
    - User sorted 6 lines of 15 colored tiles (3 at a time)
    - Colors chosen from a circle in Lab space centered on 0,0 in a,b and of uniform lightness.
- Motivation for users:
  - Users receive a color vision score and graphic to share on social media [see Figure 3]

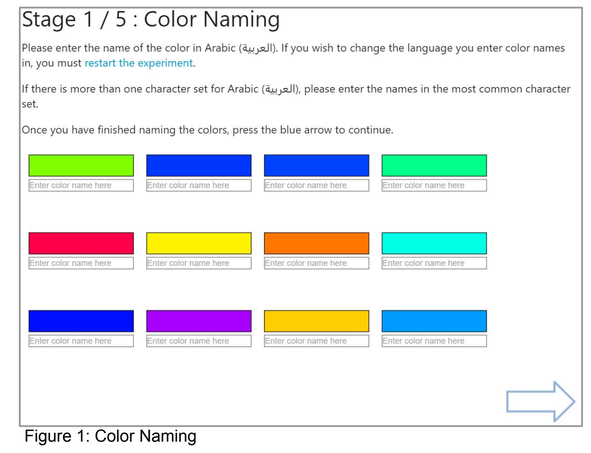


Figure 1: Color Naming

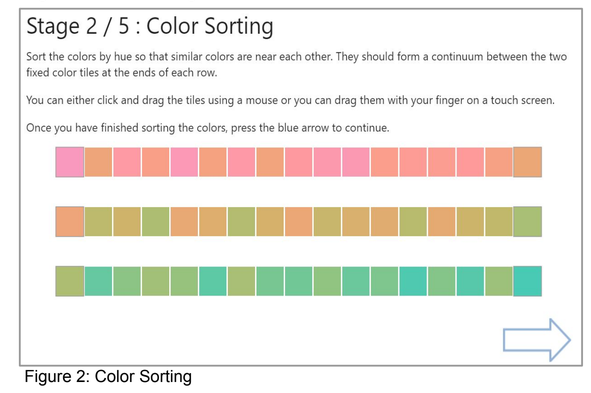


Figure 2: Color Sorting

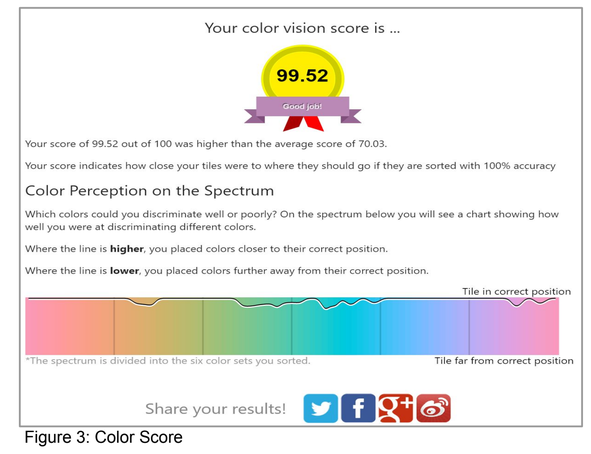
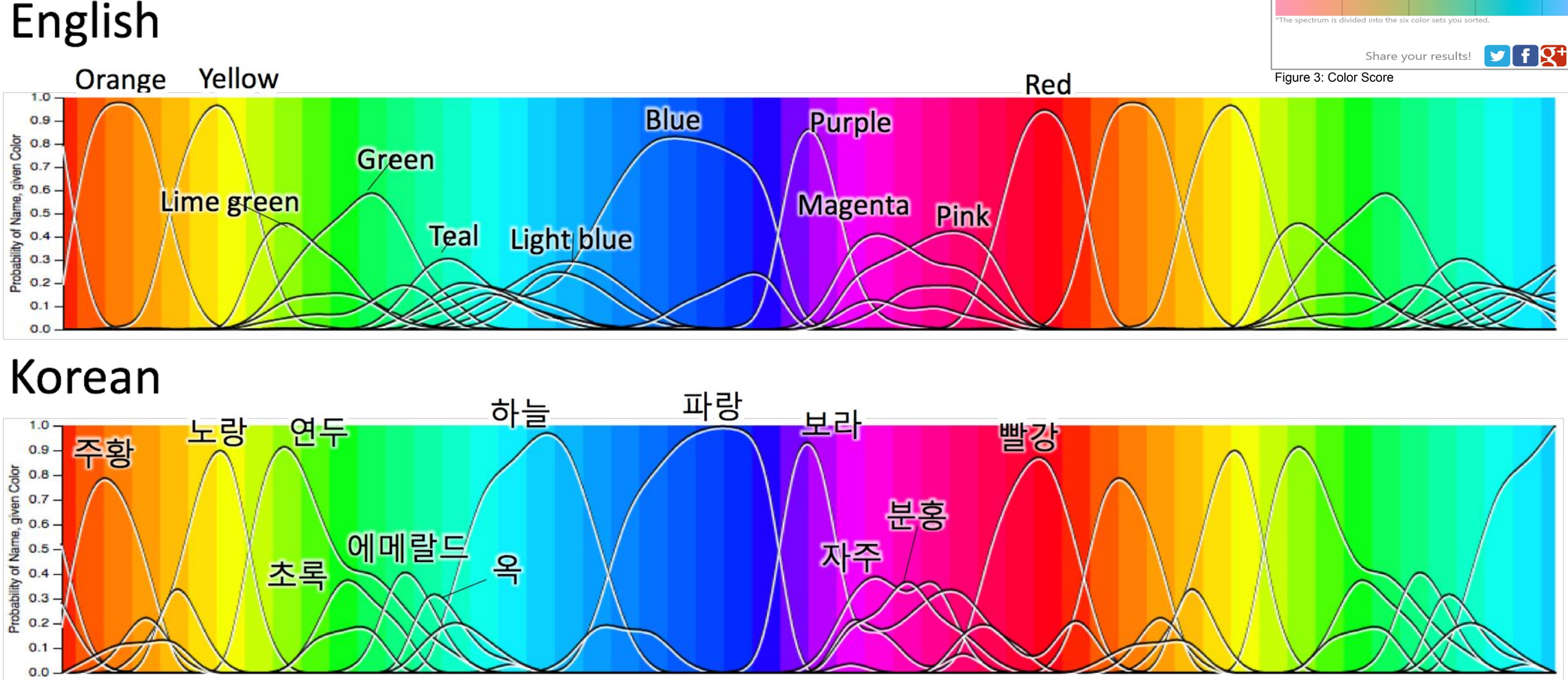


Figure 3: Color Score



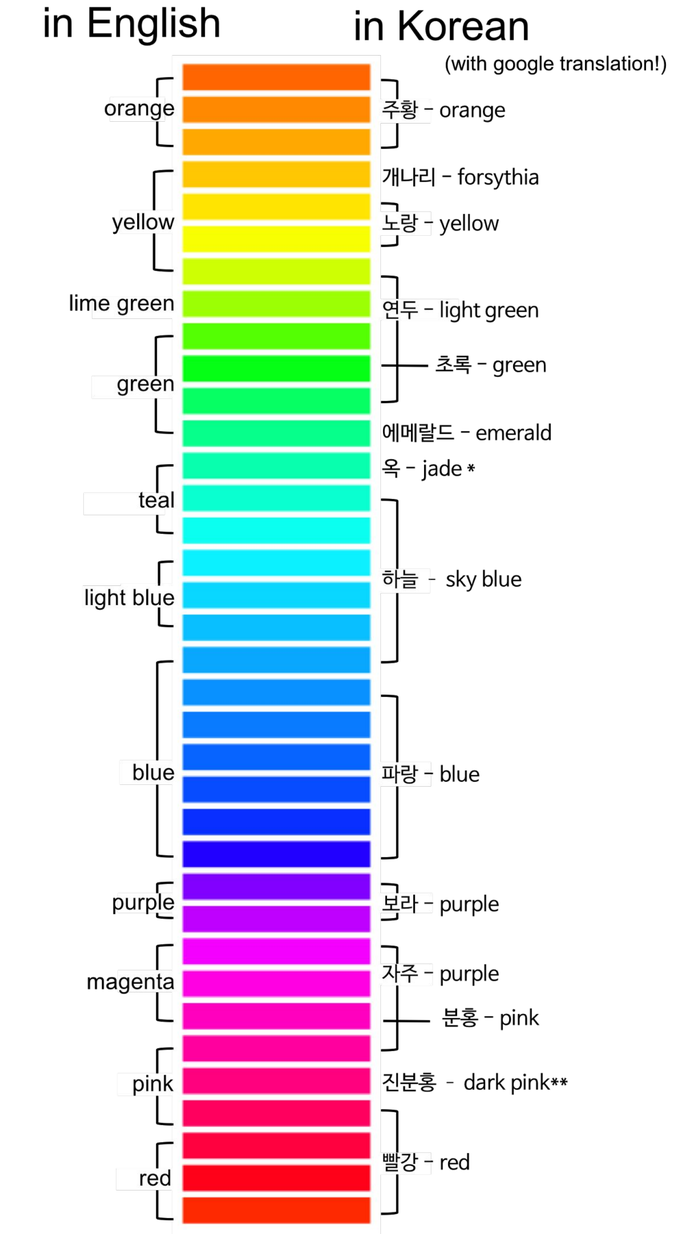
Analysis & Results

- Sparse preliminary data allowed for a limited quantitative exploration
- Data was collected from 347 participants
- More than a dozen languages were represented
  - 83.6% English, 8.9% Korean, 7.5% other, including Spanish, Chinese, Polish, and German
- There were enough Korean and English respondents to explore the pattern differences in color naming

- Colors across the Lab space were placed into 1 of 36 bins [as represented in the graphic to the right]
- Probabilities were calculated for the likelihood of a bin to be named one of 20 colors (ie “blue” or “red”)
- Probability distributions are represented in the line charts above
  - Greater values indicate a greater probability that a given bin will be given a particular color name
  - It is evident that there are some clear distinctions between English and Korean
    - Differences in the saliency of Blue-Green and Orange spectrums

- The ratio of distinct names out of total names given to a bin were calculated for Korean and a series of randomly sampled subsets of English (group N = 803)
- Paired-samples analyses revealed a significant difference between the English and Korean ratios for approximately 71% of pairwise comparisons
  - English tended to have larger ratios (more distinct words/total words applied to a bin)

- Further exploration of the more complete dataset will allow for increasingly robust conclusions
- Unsurprisingly, plots of the ratio distributions tend to mirror the general shape of the probability distributions above
  - Taken together, the two approaches begin to elucidate the nuances of color naming differences between the two focal languages



*’옥’ is translated to ‘jade’ but ‘옥색’, which means ‘jade color’, is translated to ‘green’.
**’진분홍’ is translated to ‘Jinbunhong’, which is actual pronunciation of ‘진분홍’. So we translated ‘진한 분홍’.

Limitations & Future Work

- Limited data space
  - To avoid to collect data sparsely, we limited the color space by sampling from only some edges of the rgb cube
  - We didn’t get dense responses in languages except Korean and English
  - A clear next step is collecting more broad data

- Biased demographics
  - More than a half of participants have been or are in graduate school. (28.8% graduate school, 19.9% PhD, 2.1% Postdoc)
  - Instructions were in English, so our users presumably all could read English

- Other considerations
  - Spelling and structure differences may obscure or overemphasize color naming differences
  - Respondents reported eye-strain during the task
  - Feedback indicated we might need to give more explicit directions in future releases

Citations

- Chen, Y., Kim, Y., Thayer, K., and Wang, J. (2015). Color perception in different languages. Unpublished.
- Chuang, J., Stone, M., and Hanrahan, P. (2008). A probabilistic model of the categorical association between colors. In Color Imaging Conference. 6–11. http://vis.stanford.edu/papers/color-names
- Farnsworth, D. (1957). The Farnsworth--Munsell 100 -hue test for the examination of color discrimination. Munsell Color Company.
- Reinecke, K. and Gajos, K. Z. (2015). Labinthewild: Conducting large--scale online experiments with uncompensated samples. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15, pages 1364–-1378, New York, NY, USA, 2015. ACM.
- Reinecke, K., Flatla, D. R., & Brooks, C. (2016). Enabling designers to foresee which colors users cannot see. CHI‘ 16, May 07–12, 2016, San Jose, CA, USA. DOI: http://dx.doi.org/10.1145/2858036.2858077
- Winawer, J., et al. (2007). Russian blues reveal effects of language on color discrimination. In *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences. http://www.pnas.org/content/104/19/778
- Wuerger, S. M., et al. (2012). Blue–green color categorization in Mandarin–English speakers. *Journal of the Optical Society of America*, 29 (2), 102–107. DOI: http://dx.doi.org/10.1364/JOSAA.29.