Focal Colors Review Among the WCS Language Families

Bohdan Hurskyi
1009773149

Kateryna-Mariia Blavt
1009773238

Abstract

The hypothesis is that despite the very different geographical locations of specific language families, their focal colors are close to universal focal colors across all languages in WCS. Analyzing provided data and reproducing local focal colors on a map shows that patterns of all languages that were studied are similar to each other and similar to the world map. The patterns show us that there are clustered areas of foci near "red", "green", "yellow", and "blue" ("black" and "white" we don't take into consideration). Pearson correlation values support our hypothesis as well.

Keywords: WCS; Pearson correlation; focal colors; language families

Introduction

We were inspired by the first part of the "Focal colors are universal after all" research (Regier, Kay, & Cook, 2005). In that paper, they examined color-naming data from languages of 110 nonindustrialized societies and showed that best-example choices for color terms in these languages cluster near the prototypes for English white, black, red, green, yellow, and blue. Taking this paper as the basis of our research, in our hypothesis we wanted to confirm and vividly show it on three different language families with different geographical locations. As the alternative results for the alternative hypothesis, the maps of some language families may have peculiar extreme points that differ from the expected ones. For example, the focal color chip for any color showed off the or even new unexpected focal appears.

Methodology

First of all we would like to make clear the way we have chosen our language families(Austronesian, Oto-Manguean, Gur). From the base work "The World Color Survey" (Kay, Berlin, & Maffi, 2011) we have received the data of the origin of the languages represented in the WCS. Using this data we have calculated the largest families represented in the survey, having the result we have chosen the three representatives from different continents, with more than **5 languages** each.

Here we count the 4 focal colors representing maps (one for the whole world and 3 for each of the languages). The idea behind the methodology of calculating is the reproduction of the brightness/hue table as the matrix of the same dimensions, where each element will save the number of times the corresponding chip was named as the focal. It is implemented in the **countFoci** function: we create the matrix, iterate through the **fociData** (data representing the chips called as focal for every attendant of the survey), calculate the coordinates of the chip in our matrix, increment the value of the element. Lastly, we remove the values of chips that represent "black" and "white" colors because they don't give any value to our research

To count the statistical correlation between the data we use the **Pearson correlation coefficient**. It measures the strength of the linear relationship between our matrices. You may see the results of this measurements in the Table 1 in the Results section.

To notice some interesting regularities and differences we have decided to visualize our data. In one of the references papers(Sturges & Whitfield, 1995) we have noticed the contour plot representation of the focal colors and chosen the method for our research because it represents the density of two-dimensional data pretty vividly. We have used the **Matplotlib** library to plot our data using the **axis.contour** function, where we specified the labels for every language family, since the amount of responses vary significantly among different families.

Results

Below are the statistical and visual results of our research:

Statistical analysis

Pearson's Correlation combinations:

Table 1: Correlation between the foci matrices

Foci 1	Foci 2	Correlation
World	Austronesian	0.9182
World	Oto-Manguean	0.8722
World	Gur	0.8295
Austronesian	Oto-Manguean	0.8132
Austronesian	Gur	0.7947
Oto-Manguean	Gur	0.6852

Figures

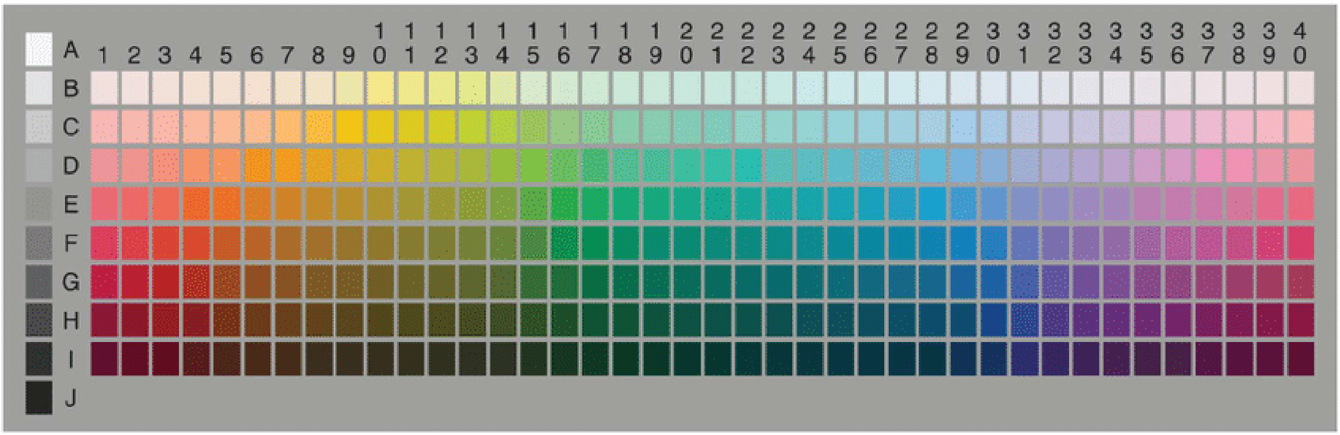

Figure 1: Munsell color chart.

Firstly we have plotted the focal color map across all the 110 languages. We have received the similar results as in the "Focal colors are universal after all", i.e. There are six universal focal colors: "black", "white", "red", "yellow", "blue", "green". Four of them are clearly seen on the Figure 2, the regions with the high values represent the popularity of the chips as the focal colors. You can do the comparison yourself using the coordinates of the Figure 1.

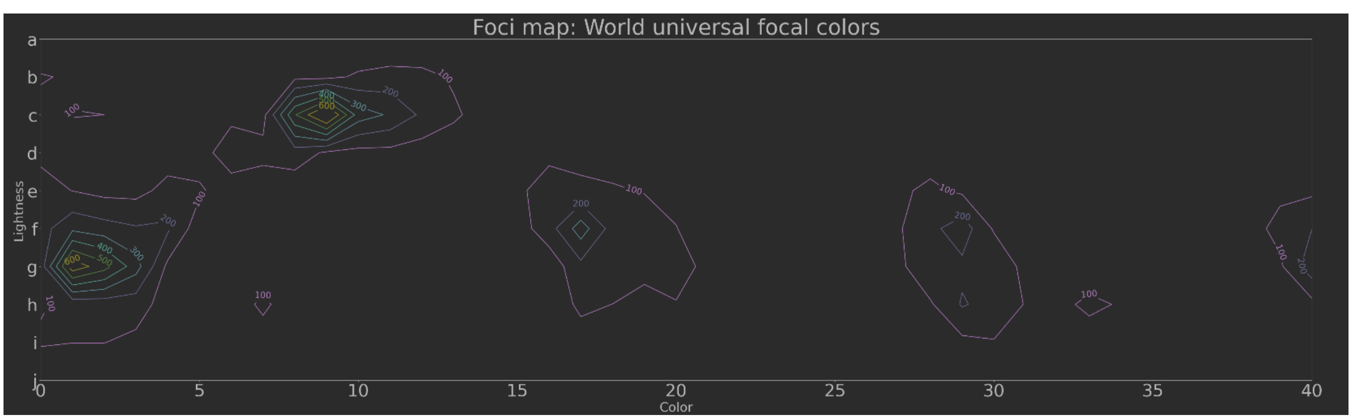

Figure 2: Contour plot of the whole world foci.

Next we have the focal color map for the Gur family. All the main focal colors from the Figure 2 are present in the Figure 3. Still the ”red” and ”yellow” focal colors are more dominant beyond other. The first noticeable difference is the region in C30-C35 coordinates, where we can see some focal tendency, in English terms we can characterize those as ”purple” colors.

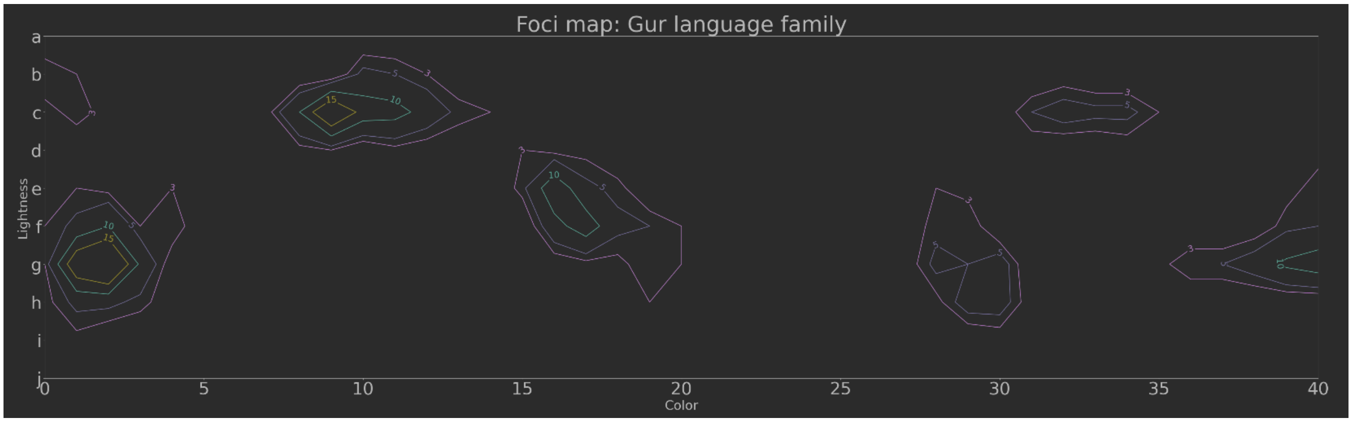

Figure 3: Contour plot of the Gur family foci.

The next one is the focal color map for the Oto-Manguean family. All the main focal colors from the Figure 2 are present in the Figure 4. Still, there is distinguish region in H5-H9 region, where we can see the ”brown” focal color on the Figure 4. Also there is an uncertainty in the ”blue” color region on Figure 4 which may refer to the differences among the languages inside the Oto-Manguean family, similar to the reference study (webster & kay, 2005).

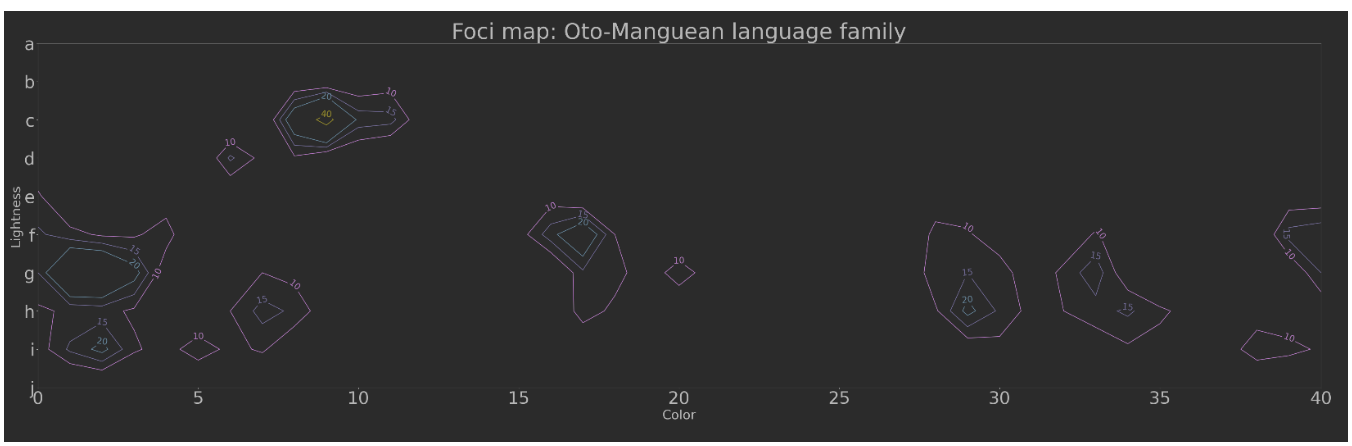

Figure 4: Contour plot of the Oto-Manguean family foci.

The last one is the focal color map for the Austronesian family. All the main focal colors from the Figure 2 are present in the Figure 5. The Figure 2 and Figure 5 have no significant differences, so correspond really well. The single noticeable exclusion is absence of the vivid peak in the ”green” region on the Figure 5, what may signal the differences among the languages in the Austronesian family, similar to the reference study (webster & kay, 2005).

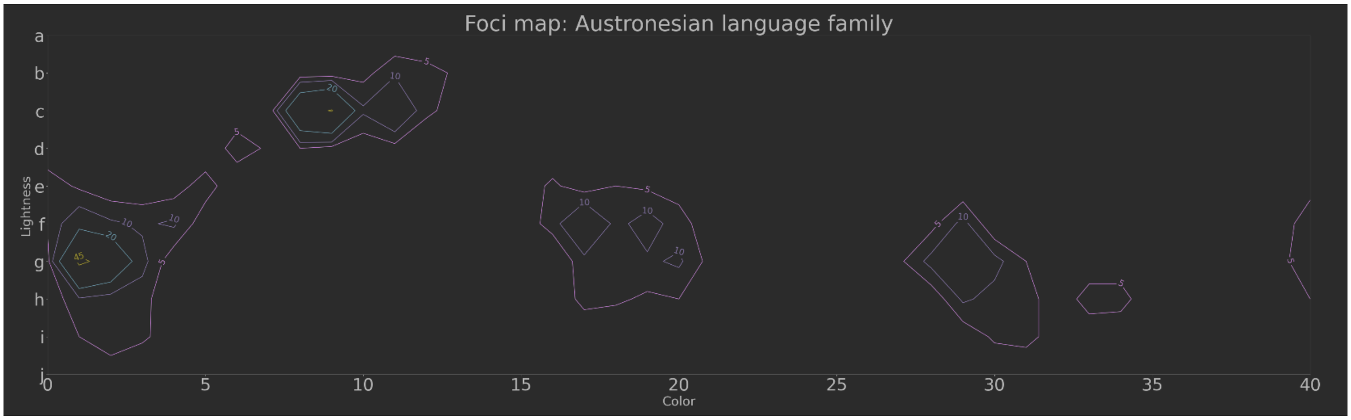

Figure 5: Contour plot of the Austronesian family foci.

Conclusion

In conclusion, we approved our hypothesis that in spite of different locations focal colors are almost the same. Yellow and red with higher concentrated areas, which means that more people choose the same chips for 'yellow' and 'red' as focal colors in their language. 'Blue' and 'green' regions are quite broad on every map as well as on the world map which means that the distribution of chosen chips is bigger. Anyway, high Pearson correlation values that are more than 0.80 between each language results and world map indicate that there is a correspondence between focal colors in different languages and chips they chose as focal colors.

To extend our research we can analyze languages with different amount of focal colors and find out the "development" factors. based on this data we may also model the way the languages and the comprehension of colors develop, similar to the reference work (Abbott, Griffiths, & Regier, 2016). Also, we can propose how different peaks of each color are based on landscape colors in each region but for this information we face the limitation of lack of data.

References

Abbott, J. T., Griffiths, T. L., & Regier, T. (2016, 09). Focal colors across languages are representative members of color categories. *Proceedings of the National Academy of Sciences*, *113*, 11178-11183. doi: 10.1073/pnas.1513298113

Kay, P., Berlin, B., & Maffi, L. (2011). *The world color survey*. Stanford Univ Center for the Study.

Regier, T., Kay, P., & Cook, R. S. (2005, 05). Focal colors are universal after all. *Proceedings of the National Academy of Sciences*, *102*, 8386-8391. doi: 10.1073/pnas.0503281102

Sturges, J., & Whitfield, T. W. A. (1995, 12). Locating basic colours in the munsell space. *Color Research Application*, *20*, 364-376. doi: 10.1002/col.5080200605

webster, m. a., & kay, p. (2005, 08). variations in color naming within and across populations. *Behavioral and Brain Sciences*, *28*, 512-513. doi: 10.1017/s0140525x05470089

1 Code Appendix

1.1 Data preparation

```
import numpy as np
import pandas as pd
import matplotlib as plt
from scipy import stats
from random import random
%matplotlib inline
from wcs_helper_functions import *

#%% md
Getting the data using helper functions
#%%
fociData = readFociData('./WCS_data_core/foci-exp.txt');
munsellInfo = readChipData('./WCS_data_core/chip.txt');
origin = pd.read_csv('./WCS_data_core/origin.txt', sep="	", header=None)
#%% md
Here we discover our language families in order to choose the best for our research
#%%
origin.columns = ["index", "lang_name", "family", "country"]
origin
#%% md
We have chosen: Oto-Manguean(Mexico), Austronesian(Philippines, Papua New Guinea), Gur(Ivory Coast, Ghana)
#%%
display(origin["family"].value_counts())
austr_list, oto_list, gur_list = list(origin[(origin["family"]=="Austronesian")]["index"]),list(origin[(origin["family"]=="Oto-Manguean")]["index"]),list(origin[(origin["family"]=="Gur")]["index"])
print(austr_list)
print(oto_list)
print(gur_list)
#%% md
The function to count the matrices of chips named as focal
#%%
def countFoci(list_of_indices):

    matr_world = np.zeros((10,41))
    for lang in list_of_indices:
        for speaker in fociData[lang]:
            for color in fociData[lang][speaker]:
                for cood in fociData[lang][speaker][color]:
                    coods = cood.split(":")
                    y,x = ord(coods[0])-65, int(coods[1])
                    matr_world[y][x] += 1
    matr_world[0][0] = 0
    matr_world[9][0] = 0
    return np.flipud(matr_world)



matr_oto = countFoci(oto_list)
matr_austr = countFoci(austr_list)
```

```
matr_gur = countFoci(gur_list)
matr_wcs = countFoci(fociData)
matr_gur
#%% md
Pierson's Correlation
#%%
import inspect

def normalize(matr):
    return np.divide(matr, sum([sum(i) for i in matr]))


def correlation_coefficient(T1, T2):
    numerator = np.mean((T1 - T1.mean()) * (T2 - T2.mean()))
    denominator = T1.std() * T2.std()
    if denominator == 0:
        return 0
    else:
        result = numerator / denominator
        return round(result,4)

def retrieve_name(var):
    callers_local_vars = inspect.currentframe().f_back.f_locals.items()
    return [var_name for var_name, var_val in callers_local_vars if
        var_val is var]
#%%
print("Correlation between the whole world data and Oto-Manguean language
     family:",correlation_coefficient(normalize(matr_wcs),normalize(
    matr_oto)))
print("Correlation between the whole world data and Austronesian language
     family:",correlation_coefficient(normalize(matr_wcs),normalize(
    matr_austr)))
print("Correlation between the whole world data and Gur language family
    :",correlation_coefficient(normalize(matr_wcs),normalize(matr_gur)))
#%% md
Some combinations of our languages to notice some interesting
    coincidences
#%%
langlist = [matr_oto,matr_austr,matr_gur]
for i, a in enumerate(langlist):
    for b in langlist[i:]:
        print(correlation_coefficient(normalize(a), normalize(b)),
            retrieve_name(a),retrieve_name(b))
```

1.2 Data visualization

```
X, Y = np.meshgrid(range(0,41), range(0,10))
fig = plt.figure(figsize=(41, 11))
ax = fig.add_axes([1,1,1,1])
cp = ax.contour(X, Y, matr_wcs, levels=[100,200,300,400,500,600])
ax.clabel(cp, inline=True, fontsize=20)
ax.set_yticklabels(["j","i","h","g","f","e","d","c","b","a"])
ax.set_title('Foci map: World universal focal colors',fontsize=50)
ax.set_xlabel('Color',fontsize = 30)
ax.set_ylabel('Lightness',fontsize = 30)
plt.xticks(fontsize=40)
```

```
plt.yticks(fontsize=40)
#%%
X, Y = np.meshgrid(range(0,41), range(0,10))
fig = plt.figure(figsize=(41, 11))
ax = fig.add_axes([1,1,1,1])

cp = ax.contour(X, Y, matr_oto, levels=[10,15,20,40])
ax.clabel(cp, inline=True, fontsize=20)
ax.set_yticklabels(["j","i","h","g","f","e","d","c","b","a"])
ax.set_title('Foci map: Oto-Manguean language family',fontsize=50)
ax.set_xlabel('Color',fontsize = 30)
ax.set_ylabel('Lightness',fontsize = 30)
plt.xticks(fontsize=40)
plt.yticks(fontsize=40)

#%%
X, Y = np.meshgrid(range(0,41), range(0,10))
fig = plt.figure(figsize=(41, 11))
ax = fig.add_axes([1,1,1,1])

cp = ax.contour(X, Y, matr_austr, levels=[5,10,20,45])
ax.clabel(cp, inline=True, fontsize=20)
ax.set_yticklabels(["j","i","h","g","f","e","d","c","b","a"])
ax.set_title('Foci map: Austronesian language family',fontsize=50)
ax.set_xlabel('Color',fontsize = 30)
ax.set_ylabel('Lightness',fontsize = 30)
plt.xticks(fontsize=40)
plt.yticks(fontsize=40)

#%%
X, Y = np.meshgrid(range(0,41), range(0,10))
fig = plt.figure(figsize=(41, 11))
ax = fig.add_axes([1,1,1,1])

cp = ax.contour(X, Y, matr_gur, levels=[3,5,10,15])
ax.clabel(cp, inline=True, fontsize=20)
ax.set_yticklabels(["j","i","h","g","f","e","d","c","b","a"])
ax.set_title('Foci map: Gur language family',fontsize=50)
ax.set_xlabel('Color',fontsize = 30)
ax.set_ylabel('Lightness',fontsize = 30)
plt.xticks(fontsize=40)
plt.yticks(fontsize=40)
```