

Color terms across Europe

An investigation of the interaction between color terms and European language families

S.A. (Simon) Claassen

Research Master *Linguistics and Communication Sciences*, Radboud University Nijmegen

Abstract: Languages show variation in the colors that they have a distinct color term for, with some languages only distinguishing two colors and some having distinct terms for up to eleven different colors. These appear to follow a hierarchy, where colors at a higher place in the hierarchy will have color terms in a certain language before those at a lower place. Although many studies have been conducted so far that focused on the different color terms found in various languages, there have as of yet been no studies that have attempted to find patterns in the distributions of these terms across related languages. Therefore, this paper investigated what the etymological origins of the color terms of European languages are. Additionally, it examined whether the forms of color terms found in languages are categorically more similar when these languages are from the same language family or geographical area. Finally, it investigated whether the number of colors distinguished within a language can be related to the language family this language belongs to. For this purpose, the color terms of fifty European languages, some related and some unrelated, were collected and compared. Firstly, it was found that languages usually acquire color terms by inheriting them from their proto-language, especially for the upper colors in the hierarchy, whereas other strategies to acquire color terms, such as borrowing or morphological derivation, are generally only used for the lower colors of the hierarchy. This supports the aforementioned color hierarchy, as it implies that the upper colors of the hierarchy are more semantically stable across different stages of language development. It was also found that the color terms found in a language are influenced to a greater degree by its language family than its geographical location. Finally, no link was found between the numbers of color terms found in closely related languages.

Key words: color terms, color hierarchy, language family, Indo-European linguistics, historical linguistics

1. Introduction

As human beings, we constantly experience the world around us through our senses. Arguably our most important sense, vision, allows us to perceive objects in our surroundings through the light reflected by these objects (Livingstone, 2002). The wavelengths of the electromagnetic spectrum that are reflected or absorbed by objects determine the colors these objects appear to have to our eyes. The color spectrum of visible light encompasses wavelengths ranging from approximately 400 to 700 nm, with the lower wavelengths of the spectrum appearing to our eye as violet and the higher ones appearing as red (Buser & Imbert, 1992).

Although the color spectrum forms a continuous unit comprising numerous different hues gradually transitioning from one to another, languages require people to categorically divide the spectrum into distinct colors for the sake of naming these colors with respective color terms. Languages have been shown to display considerable variation with regard to the colors they distinguish (Kay, Berlin, Maffi, & Merrifield, 1997). Most European languages linguistically distinguish a wide variety of colors and specific shades, whereas there are also languages, such as Tsimané, that reportedly have distinct terms for only two or three colors (Gibson et al., 2017), with additional claims that there are even languages without any color terms, such as Pirahã (Regier, Kay, & Khetarpal, 2009).

Generally, however, languages do not appear to distinguish more than eleven distinct color categories. These color terms appear to follow a hierarchy, where colors higher up in the hierarchy must be present in languages before colors further down in the hierarchy (Berlin & Kay, 1991). This color hierarchy is shown in Figure 1 below. The semantic boundaries of color terms also appear to vary from language to language. For instance, whereas most European languages have two separate words for the colors green and blue, many languages group these two colors together under a single term (Davies & Corbett, 1997).

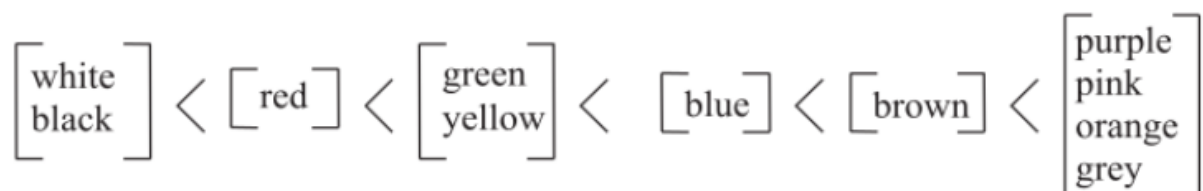


Figure 1. The color hierarchy (Berlin & Kay, 1991).

These observations have raised several questions related to cognition. One could wonder, for instance, whether the color terminology of a language shapes or influences its speakers' perception of certain colors or whether people are born with certain color

concepts already in place. It appears that language does indeed influence color perception to some degree, but there also appears to be evidence for the existence of universally present color categories, which are often stated to be black, white, red, green, yellow and blue (Regier & Kay, 2009).

Over the years, there have been multiple studies that attempted to report on the different color terms of numerous languages. The biggest study to do so was the World Color Survey, which gathered data on color terminology from 110 languages (Kay, Berlin, Maffi, Merrifield, & Cook, 2009). From studies such as this one, it becomes evident that the languages of the world are very diverse with regard to the form, meaning and categorization of their color terms. Languages can have both morphologically simple and complex forms for color terms, have just a few or quite a lot of color terms, distinguish certain colors or group multiple colors under a single term and place the boundaries between certain colors in different places.

It would be interesting to relate the dispersion of color terms across the languages of the world to certain topics in linguistic typology and historical linguistics. The continent of Europe is a typologically interesting geographical area. Almost all of the languages spoken on the continent, both national and regional, belong to the same linguistic family, the Indo-European languages. Languages belonging to the same subfamily within this language family can be clearly seen to share more features with each other than with other branches within the family, quite often because certain words and grammatical structures in both languages derive from their common ancestor language, often called a proto-language (Renfrew, 1989).

One can thus imagine that the color terms found in languages and, by extension, the demarcations of the color categories these color terms refer to might be more similar to each other the more closely related these languages are. Another factor could be geographical distribution (Epps & Huehnergard, 2013). As languages that are spoken in the same geographical area and have been in contact with each other over time tend to adopt words from each other, it is also possible that languages show similarities in their color terms with the languages in their direct vicinity.

Analyzing a sample of fifty European languages, I will examine three aspects of the relations of color terminology between these languages, corresponding to my three main research questions. I will investigate what the etymological origins of the color terms of European languages are, whether the forms of color terms found in languages are categorically more similar when these languages are from the same language family or geographical area and whether the number of colors distinguished within a language can be related to the language family this language belongs to.

As for the first question, I expect that most color terms for colors high up in the color hierarchy cannot be traced back to etymologically originate in a so-called source-based term, which is a color term that is clearly named after an object of this color (de Valk, Wnuk, Huisman, & Majid, 2017). Additionally, I expect that some color terms, especially those for the lower colors of the color hierarchy, will have a clearly source-based origin or be derived from other color terms. I also expect that color terms that cannot be traced back to originate in source-based terms will mostly, though not necessarily always, retain the same semantic meaning and thus refer to the same color throughout different proto-stages of languages.

As for the second question, I hypothesize that languages within the same families are more likely to have the same color terms and that languages within the same geographical area are also more likely to share certain color terms, although presumably to a lesser degree than languages within the same family. Finally, as for the final question, I hypothesize that there will be few differences in the number of colors distinguished by European languages, as I expect that most European languages distinguish all of the eleven colors in the sample, or else ten or nine at least, and that, if there are differences in the number of colors distinguished, these will likely pattern with language families.

2. Method

In order to test the hypotheses mentioned above, I collected language data on color terms from a sample of fifty European languages to form a relatively representative sample. To this end, the selected languages are roughly equally dispersed across Europe geographically and across the various language families and subfamilies of Europe. The majority of languages in the sample are Indo-European languages belonging to one of the six branches of this family spoken in Europe, along with three more languages from the Uralic family as well as the Basque and Maltese languages.

All languages in the sample are proportionally distributed across the primary and secondary levels of their respective language families. The three biggest subgroups in the sample, the Germanic, Italic and Balto-Slavic languages, are further divided into the West and North Germanic, the Western, Central, Southern and Eastern Romance and the West, South and East Slavic and Baltic subfamilies, respectively. Of course, many of these language families can be further divided into even smaller subgroups, but for practical reasons, this classification was used. The languages included in the

sample are shown in Table 1 below. In Figure 2, one can see the geographical distribution of these languages throughout Europe.

Table 1

Languages included in the sample and the (sub)families they belong to.

Language	Family	Language	Family
English	Indo-European; Germanic (West)	Manx	Indo-European; Celtic
Frisian	Indo-European; Germanic (West)	Irish	Indo-European; Celtic
Dutch	Indo-European; Germanic (West)	Gaelic	Indo-European; Celtic
German	Indo-European; Germanic (West)	Welsh	Indo-European; Celtic
Danish	Indo-European; Germanic (North)	Cornish	Indo-European; Celtic
Swedish	Indo-European; Germanic (North)	Breton	Indo-European; Celtic
Norwegian	Indo-European; Germanic (North)	Greek	Indo-European; Hellenic
Icelandic	Indo-European; Germanic (North)	Albanian	Indo-European; Albanian
Faroese	Indo-European; Germanic (North)	Polish	Indo-European; Balto-Slavic (West)
Romansh	Indo-European; Italic (West)	Czech	Indo-European; Balto-Slavic (West)
Ladin	Indo-European; Italic (West)	Slovak	Indo-European; Balto-Slavic (West)
Friulian	Indo-European; Italic (West)	Slovene	Indo-European; Balto-Slavic (South)
Lombard	Indo-European; Italic (West)	Serbian	Indo-European; Balto-Slavic (South)
Piedmontese	Indo-European; Italic (West)	Bulgarian	Indo-European; Balto-Slavic (South)
French	Indo-European; Italic (West)	Macedonian	Indo-European; Balto-Slavic (South)
Occitan	Indo-European; Italic (West)	Russian	Indo-European; Balto-Slavic (East)
Catalan	Indo-European; Italic (West)	Ukrainian	Indo-European; Balto-Slavic (East)
Aragonese	Indo-European; Italic (West)	Belarusian	Indo-European; Balto-Slavic (East)
Asturian	Indo-European; Italic (West)	Lithuanian	Indo-European; Balto-Slavic (Baltic)
Spanish	Indo-European; Italic (West)	Latvian	Indo-European; Balto-Slavic (Baltic)
Portuguese	Indo-European; Italic (West)	Estonian	Uralic; Finnic
Italian	Indo-European; Italic (Central)	Finnish	Uralic; Finnic
Sicilian	Indo-European; Italic (Central)	Hungarian	Uralic; Ugric

Sardinian	Indo-European; Italic (South)	Basque	Vasconic
Romanian	Indo-European; Italic (East)	Maltese	Afroasiatic; Semitic



Figure 2. Geographical distribution of languages included in the sample.

For each of these languages, appropriate linguistic descriptions or lexicons were collected and terms for the colors red, orange, yellow, green, blue, purple, pink, white, gray, black and brown were retrieved from these sources. Usually, it was quite straightforward and easy to see what the main term used for a color was, but if there were multiple terms in common usage for a certain color in a language, the term that appeared to be more frequent or more basic in meaning or usage was chosen. This, however, was not always obvious. For instance, the German word *purpurn* appears to be a pretty basic term for the color purple, but it is not really used by German speakers anymore, with *violett* being the preferred term.

3. Results

The sample of color terms retrieved from the literature can be found in the Appendix, as well as the sources used for each language. As can be seen, each of the fifty languages in the sample has a color term of some sort for each of the eleven color categories. For the six highest colors of the color hierarchy, black, white, red, yellow, green and blue, many languages have a color term they inherited from their respective proto-language. Most of them are monomorphemic and cannot be traced back to have source-based origins.

Almost all of the terms for these colors also retained their original meaning and thus consistently referred to the same color through different stages of language development. For instance, the Dutch word *rood* “red” descends from Proto-Germanic **raudaz* “red”, which in turn descends from the Proto-Indo-European root **h₁rewdʰ-* “red” (Kroonen, 2013). Similarly, the Spanish word *verde* “green” descends from Proto-Romance/Vulgar Latin *viridis* “green” and ultimately from the Latin word *viridis* “green” (de Vaan, 2008). As can be seen, these two terms were passed down from two proto-languages to their descendants and retained their meaning over time.

There were only a few instances where the most commonly used term in a language for one of the six highest colors of the hierarchy originated in a source-based term or a term for a different color. One of the most notable examples is the Slovene word *rumena* “yellow”, which derives from Proto-Slavic **ruměnbъ* “reddish, rosy”, also from Proto-Indo-European **h₁rewdʰ-* “red” (Derksen, 2007). Slovene does have another word for yellow, *žolta*, which is a cognate of the word for yellow in the other Slavic languages, but this word appears to be very rare, apparently having been largely displaced by *rumena*.

Another interesting example of such a semantic shift concerns Lithuanian *mėlyna* “blue” and Latvian *melns* “black”, which both derive from Proto-Baltic **mel(n)-* “black, blue” and ultimately from Proto-Indo-European **mel-* “dark, red” (Derksen, 2014). As can be seen, these two terms for colors at the top of the hierarchy both derive from a term originally referring to a different color, possibly through the meaning of “dark” of this latter term. Not only have these two color terms diverged in meaning from their origin, they have also diverged from each other over time.

If languages lack monomorphemic or etymologically stable terms for certain colors, this happens almost exclusively in the lower colors of the hierarchy. Languages occasionally derive some of these colors directly from other color terms present in the language. For instance, the Danish word *lyserød* “pink” and the Faroese word *ljósareyður* “pink” are both derived from the respective words for red in these languages, both literally translating to “light red”. Similarly, the Manx words *jiarg-*

bwee “orange”, *jiarg-gorrym* “purple” and *jiarg-bane* “pink” literally translate to “red-yellow”, “red-blue” and “red-white”, respectively.

Languages also appear to resort to source-based terms increasingly for the lower colors of the hierarchy, although the boundary between actual source-based terms and lexicalized color terms has proven to be somewhat vague. Clear examples of source-based terms are the Icelandic and Faroese words *appelsínugulur* “orange” and *brandgulur* “orange”, literally meaning “orange yellow”, where “orange” refers to the fruit rather than the color, and “fire yellow”, respectively.

Table 2

Numbers of shared etymologies within language families.

Language family												
Germanic (9)	9	8/1	9	5/4	9	9	9	7/1/1	3/3/2/1	4/2/1/1/1	7/2	
West Germanic (4)	4	3/1	4	4	4	4	4	4	2/1/1	3/1	2/2	
North Germanic (5)	5	5	5	5	5	5	5	3/1/1	3/2	2/1/1/1	5	
Italic (16)	14/2	15/1	8/4/2/2	10/4/2	16	11/4/1	11/2/1/1/1	14/1/1	9/3/2/1/1	16	15/1	
Western Romance (12)	11/1	11/1	5/3/2/2	7/4/1	12	8/4	8/2/1/1	11/1	6/3/2/1	12	12	
Central Romance (2)	2	2	2	2	2	2	2	2	2	2	2	
Southern Romance (1)	1	1	1	1	1	1	1	1	1	1	1	
Eastern Romance (1)	1	1	1	1	1	1	1	1	1	1	1	
Celtic (6)	3/3	6	3/2/1	3/3	3/3	3/3	3/2/1	4/1/1	4/1/1	2/2/1/1	5/1	
Hellenic (1)	1	1	1	1	1	1	1	1	1	1	1	
Albanian (1)	1	1	1	1	1	1	1	1	1	1	1	
Balto-Slavic (12)	12	10/1/1	8/1/1/1/1	11/1	12	5/4/1/1/1	3/2/2/2/1/1/1	9/1/1/1	6/5/1	12	4/4/2/2	
West Slavic (3)	3	3	3	3	3	2/1	2/1	2/1	2/1	3	2/1	
South Slavic (4)	4	4	3/1	3/1	4	2/2	2/1/1	2/1/1	2/1/1	4	4	
East Slavic (3)	3	3	2/1	3	3	3	3	3	3	3	3	
Baltic (2)	2	2	1/1	2	2	1/1	1/1	2	1/1	2	2	
Uralic (3)	2/1	2/1	2/1	2/1	1/1/1	2/1	2/1	3	2/1	2/1	1/1/1	
Finnic (2)	2	2	2	2	1/1	2	1/1	2	2	1/1	1/1	
Ugric (1)	1	1	1	1	1	1	1	1	1	1	1	
Vasconic (1)	1	1	1	1	1	1	1	1	1	1	1	
Afro-Asiatic (1)	1	1	1	1	1	1	1	1	1	1	1	
Semitic (1)	1	1	1	1	1	1	1	1	1	1	1	

As for the similarities between color terms within language families, Table 2 above gives an overview of the origins of the different color terms within certain language families. The number in brackets behind each family is the number of languages in the

sample within this family. The numbers in the cells indicate how many languages of a certain family share a color term with similar origins. For instance, the yellow column contains “5/4” for the Germanic family. This means that five of the nine Germanic color terms in the sample have the same etymology, while the remaining four also share an etymology. Of course, it would be more informative to perform statistical tests with these data, but as this paper is the first of its kind and its aim is largely explorative, statistics are beyond the scope of this paper.

As can be seen from the table, the homogeneity of color terms within language families varies. The words for the colors green, white and black quite often appear to have etymologies shared by all languages within a language family, whereas for other colors, such as brown, the opposite is true. For some languages, such as Greek and Albanian, this table is not very informative, of course, as they are not closely related to any other languages in the sample. For these languages, it is more informative to note that many of their color terms are etymologically unique within the sample and do not have cognates within other subfamilies. Additionally, Albanian *verdhë* “yellow” and *blerë* “green” are unique forms for the colors they refer to.

When looking at some of the language families in the sample, one can see that the languages within the family pattern together at an even lower level of classification that was not even included in the analysis. The subdivision of the Celtic languages into the Goidelic languages of Manx, Irish and Gaelic and the Brittonic languages of Welsh, Cornish and Breton, for instance, was not included in the analysis. One can see, however, that the words for the colors red, yellow, green, blue, white and brown of the Goidelic languages have the same etymological origins, as do the words for the colors yellow, green, blue, white and black of the Brittonic languages (Matasovic, 2008).

Although to a much lesser degree than its language family, the geographical location of a language can also determine the color terms it has. For instance, even though Frisian is much more closely related to English than it is to Dutch, its color terms are highly similar to those of Dutch, to the point where their forms for the color purple, which have the same etymology, are unique among the languages of Europe. Similarly, the Breton words for pink and gray are clearly borrowed from French due to their geographical proximity. Additionally, the Lithuanian, Latvian and Estonian words for yellow all originate in Proto-Balto-Slavic **gilʔtos* “species of moss” (Derksen, 2014).

Two languages with a high number of clearly borrowed color terms are Basque and Maltese. These two languages are completely unrelated to, but through geographical proximity heavily influenced by Spanish and Italian, respectively.

Interestingly, both languages have native color terms for the upper six colors of the color hierarchy and borrowed terms for the lower five, with the exception of Basque *berde* “green”, which is also borrowed. In Maltese, the borrowed color terms, which do not inflect, are in heavy contrast with the native color terms, which decline according to Arabic nonconcatenative morphology. For instance, *isfar* “yellow” becomes *saфра* when feminine and *sofor* when plural (Simpson, 2009).

After closely considering the forms and etymologies of color terms across Europe, one can also compare the color systems of languages as a whole and see if different languages in the sample distinguish different numbers of colors. When examining the sample, it appears that all languages distinguish the same eleven colors. One could argue, however, that Danish, Faroese, Manx and Irish, for instance, do not distinguish pink as a separate color due to their lack of a basic term for this color. Still, these lexical gaps appear to be very rare in the sample and do not hint at clear crosslinguistic patterns.

Technically, there are also languages within the sample that distinguish more than eleven colors. For instance, both Greek and Russian have separate terms for light blue, *galázio* and *golubój*, respectively, and Hungarian has a separate term for a darker shade of red, *vörös*. It appears that Russian speakers actually perceive light and dark blue as two different colors as a result of this (Winawer et al., 2007). Although some of these languages, especially Russian, can arguably be considered to distinguish more than eleven colors, this additional distinction does apparently not occur in related languages, such as the other Slavic languages besides Russian.

One case where a distinction of additional colors occurs in multiple languages across a language family concerns the languages Manx, Irish and Gaelic. For the color green, Manx has the word *geayney* and Irish and Gaelic have *uaine*, but these terms appear to be mostly reserved for dyed, artificially colored or light green objects. For naturally green things, such as grass or trees, Manx uses the word *glass* and Irish and Gaelic use *glas*, having the same etymology as the word for blue in the other Celtic languages. Although it remains disputable whether speakers of these languages perceive different shades of green as categorically different colors, this distinction forms an interesting pattern within the Celtic language family.

4. Discussion

In the previous, an analysis was provided for the origins of color terms across the languages of Europe, focusing on the etymologies of color terms themselves as well as a comparison of the color terms and color categories of languages within language

families. Centered around three corresponding research questions, this paper analyzed a sample of color terms gathered from fifty European languages. To a large extent, the hypotheses were confirmed. Most notably, for many of the languages in the sample, the terms for colors higher up in the color hierarchy were inherited from their respective proto-language, remained semantically stable throughout the years and were often shared between members of a language family.

As can be seen from the sample, languages can have several options to acquire color terms. They can inherit a term referring to the same color from their respective proto-language, such as Dutch *rood* “red”, inherit a term previously referring to a different color, such as Lithuanian *mėlyna* “blue”, morphologically derive a term from another term already present in the language, such as Danish *lyserød* “pink”, use a source-based term, such as Faroese *brandgulur* “orange” or, finally, borrow a term from a neighboring language, such as Basque *more* “purple”.

When looking at the origins of terms for the upper colors of the hierarchy, languages appear to mostly inherit them. Conversely, whenever languages employ one of the other strategies, it is usually in order to acquire a term for a color lower in the hierarchy. These tendencies appear to verify the suggested universality of the color hierarchy, as the semantic stability and inheritance from proto-languages seen with the terms for the upper colors of the hierarchy suggests that these colors have been consistently codified in a language throughout different stages of its development, which, according to the color hierarchy, is a prerequisite for the lower colors of the hierarchy to arise within a language.

These findings appear to be related to the findings that certain color terms, such as those for green, white and black, are very homogeneous within language families. For instance, the word for green has the same etymology for every single Italic language in the sample, because each of these languages acquired the term by inheriting it from their proto-language as a semantically stable word for green.

Conversely, the word for brown seems to have numerous different etymologies even among closely related languages, probably because languages employ a variety of strategies to acquire terms for this color, such as borrowing or using source-based terms. Of course, these observations do not apply to all language families in the sample, as among the Uralic languages, for instance, the word for green is not very homogeneous and, for example, the word for brown instead has the same etymology in all Germanic languages.

Generally, the hypothesis that more closely related languages will have similar color terms is also supported by the results. Although the homogeneity of color terms within language families varies greatly for different colors and from family to family,

it can be safely stated that the family a language belongs to, as was expected, plays a much bigger role in determining the color terms of this language than the geographical location of the language, as only a few languages in the sample display color terms that were borrowed from languages in the area.

As was also expected, all languages in the sample had a term of some sorts for the eleven colors taken into consideration here. Depending on one's definition, there might be a few languages that distinguish fewer colors, such as Danish and Manx, and there might be a few that distinguish more, such as Russian and Hungarian, but these lacking or additional distinctions appear to be an exception rather than a rule. As a result, there are, against expectations, no observable patterns of certain lacking or additional color distinctions shared between closely related languages, except maybe for the distinction of different hues of green in some Celtic languages.

One limitation to the current study concerns the sample size. Due to practical reasons, a limited selection of European languages had to be chosen, but there are of course many more distinct languages and dialects spoken all across Europe. Despite the limited sample size, however, the sample was compiled in such a way that it would proportionally represent all different language families and subfamilies spoken in Europe as well as cover the continent geographically, in order to maximize the validity of conclusions based on the sample.

Another limitation concerns the classification of the languages in the sample into subfamilies. European languages can generally be classified down to considerably low levels, with each level comprising several smaller and increasingly specific subbranches. The classification as it was used in this study was chosen for practical reasons and might be somewhat arbitrary in some ways. The effects of this on the results can be seen with, for instance, the Celtic languages, which share color terms down to a level that was not included in the analysis.

Sometimes, the retrieval of the appropriate color terms from the languages could also provide some difficulties. For instance, some languages in the sample had multiple words for certain colors. In these cases, the most basic or most frequently used term would be selected. However, since there were no concrete criteria, through which the appropriate term could be selected in these cases, this is arguably slightly problematic, especially considering that the selection of one term over another could greatly influence the proportions of shared etymologies shown in Table 2.

One might also argue that the approach that was taken in this study to search the literature for color terms is not completely sound. One could say, for instance, that it is not right to select the colors to search for in the literature beforehand, as this could bias the results towards more homogeneous findings. This concern would be very

valid if the sample consisted of highly diverse languages from all over the world with great differences in lexicon and, as a result, color terminology. As this sample consisted of European languages only, however, the expectation was that all languages in the sample would definitely have a term of some sort for the eleven colors searched for, as was stated before, which also proved to be the case.

This study is the first to directly relate color terminology to relationships between languages, so there is a lot of space for improvement and expansion in the form of follow-up studies. For instance, the numbers and proportions in Table 2 are to a certain degree informative, but it would be more interesting to see, as was stated before, if one could apply statistical analyses to such numbers, so that more definitive conclusions could be drawn from them.

Another possible follow-up study could be a comparative study of the processing of the different words for green by speakers of different Celtic languages to see whether they perceive these different shades of green as categorically different colors. This study could show whether or not discrepancies in the number of distinguished colors in a language can persist across members of the same language family.

5. Conclusion

In conclusion, this study attempted to combine several areas of research by relating previous crosslinguistic research on color terminology to language family classifications within historical linguistics. The study seems to offer additional evidence for the truthfulness and universality of the color hierarchy, as the upper colors of the hierarchy display semantically stable inheritance from proto-languages and more homogeneity across closely related languages, whereas the lower colors are often subject to borrowing from language to language and less homogeneity across languages.

Aside from this, the study hopefully also offers some promising results in general that are valuable to the research area dealing with color terminology and lays a foundation for follow-up research. By combining several different areas of research, this study hopes to contribute to research on color terminology and subsequently offer more insight into the subtle mechanics that aid humans in mapping the visual perception of color onto language.

References

- Abondolo, D. (2006a). Finnish. In *The Uralic Languages* (pp. 149-183). Abingdon-on-Thames: Routledge.
- Abondolo, D. (2006b). Hungarian. In *The Uralic Languages* (pp. 428-456). Abingdon-on-Thames: Routledge.
- Andersson, E. (2002). Swedish. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 271-312). Abingdon-on-Thames: Routledge.
- Askedal, J. O. (2002). Norwegian. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 219-270). Abingdon-on-Thames: Routledge.
- Balode, L., & Holvoet, A. (2001a). The Latvian language and its dialects. In Ö. Dahl & M. Koptjevskaja-Tamm (Eds.), *Circum-Baltic Languages* (pp. 3-40). Amsterdam: John Benjamins Publishing.
- Balode, L., & Holvoet, A. (2001b). The Lithuanian language and its dialects. In Ö. Dahl & M. Koptjevskaja-Tamm (Eds.), *Circum-Baltic Languages* (pp. 41-80). Amsterdam: John Benjamins Publishing.
- Barnes, M. P., & Weyhe, E. (2002). Faroese. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 190-218). Abingdon-on-Thames: Routledge.
- Benincà, P., Parry, M., & Pescarini, D. (2016). The dialects of northern Italy. In A. Ledgeway & M. Maiden (Eds.), *The Oxford guide to the Romance languages* (pp. 185-205). Oxford: Oxford University Press.
- Berlin, B., & Kay, P. (1991). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
- Broderick, G. (2002). Manx. In M. Ball & J. Fife (Eds.), *The Celtic Languages* (pp. 228-286). Abingdon-on-Thames: Routledge.
- Brown, W. (2002). Serbo-Croat. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 306-387). Abingdon-on-Thames: Routledge.
- Buser, P., & Imbert, M. (1992). *Vision*. Cambridge: Bradford.
- Davies, I. R., & Corbett, G. G. (1997). A cross-cultural study of colour grouping: Evidence for weak linguistic relativity. *British journal of Psychology*, 88(3), 493-517.
- de Schutter, G. (2002). Dutch. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 439-477). Abingdon-on-Thames: Routledge.
- de Vaan, M. A. C. (2008). *Etymological dictionary of Latin and the other Italic languages*. Leiden: Brill.
- de Valk, J. M., Wnuk, E., Huisman, J. L., & Majid, A. (2017). Odor-color associations differ with verbal descriptors for odors: a comparison of three linguistically diverse groups. *Psychonomic bulletin & review*, 24(4), 1171-1179.

- Demiraj, S. (2015). Albanian. In A. Giacalone Ramat & P. Ramat (Eds.), *The Indo-European Languages* (pp. 480-501). Abingdon-on-Thames: Routledge.
- Derksen, R. (2007). *Etymological dictionary of the Slavic inherited lexicon*. Leiden: Brill.
- Derksen, R. (2014). *Etymological dictionary of the Baltic inherited lexicon*. Leiden: Brill.
- Eisenberg, P. (2002). German. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 349-387). Abingdon-on-Thames: Routledge.
- Epps, P., & Huehnergard, J. (2013). Introduction: Contact among genetically related languages. *Journal of language contact*, 6(2), 209-219.
- Friedman, V. A. (2002). Macedonian. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 249-305). Abingdon-on-Thames: Routledge.
- George, K. (2002). Cornish. In M. Ball & J. Fife (Eds.), *The Celtic Languages* (pp. 410-468). Abingdon-on-Thames: Routledge.
- Gibson, E., Futrell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., Gibson, M., Piantadosi, S. T., & Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*, 114(40), 10785-10790.
- Gillies, W. (2002). Scottish Gaelic. In M. Ball & J. Fife (Eds.), *The Celtic Languages* (pp. 145-227). Abingdon-on-Thames: Routledge.
- Haberland, H. (2002). Danish. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 313-348). Abingdon-on-Thames: Routledge.
- Haiman, J. (1988). Rhaeto-Romance. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 351-390). Abingdon-on-Thames: Routledge.
- Harris, M. (1988). French. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 209-245). Abingdon-on-Thames: Routledge.
- Hoekstra, J., & Tiersma, P. M. (2002). Frisian. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 505-531). Abingdon-on-Thames: Routledge.
- Hoenigswald, H. M. (2015). Greek. In A. Giacalone Ramat & P. Ramat (Eds.), *The Indo-European Languages* (pp. 228-260). Abingdon-on-Thames: Routledge.
- Jones, M. (1988). Sardinian. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 314-350). Abingdon-on-Thames: Routledge.
- Kay, P., Berlin, B., Maffi, L., & Merrifield, W. R. (1997). Color naming across languages. *Color categories in thought and language*, 21, 58.
- Kay, P., Berlin, B., Maffi, L., Merrifield, W. R., & Cook, R. (2009). *The world color survey*. Stanford: CSLI Publications.
- Kaye, A. S., & Rosenhouse, J. (1997). Arabic Dialects and Maltese. In R. Hetzron (Ed.), *The Semitic Languages* (pp. 263-311). Abingdon-on-Thames: Routledge.

- König, E. (2002). English. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 532-565). Abingdon-on-Thames: Routledge.
- Kroonen, G. (2013). *Etymological dictionary of Proto-Germanic*. Leiden: Brill.
- Ledgeway, A. (2016). The dialects of southern Italy. In A. Ledgeway & M. Maiden (Eds.), *The Oxford guide to the Romance languages* (pp. 246-269). Oxford: Oxford University Press.
- Livingstone, M. (2002). *Vision and Art: The Biology of Seeing*. New York: Abrams.
- Mac Eoin, G. (2002). Irish. In M. Ball & J. Fife (Eds.), *The Celtic Languages* (pp. 101-144). Abingdon-on-Thames: Routledge.
- Mallinson, G. (1988). Rumanian. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 391-419). Abingdon-on-Thames: Routledge.
- Matasovic, R. (2008). *Etymological dictionary of Proto-Celtic*. Leiden: Brill.
- Mayo, P. (2002). Belorussian. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 887-946). Abingdon-on-Thames: Routledge.
- Parkinson, S. (1988). Portuguese. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 131-169). Abingdon-on-Thames: Routledge.
- Priestly, T. M. S. (2002). Slovene. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 388-454). Abingdon-on-Thames: Routledge.
- Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, 13(10), 439-446.
- Regier, T., Kay, P., & Khetarpal, N. (2009). Color naming and the shape of color space. *Language*, 85(4), 884-892.
- Renfrew, C. (1989). The origins of Indo-European languages. *Scientific American*, 261(4), 106-115.
- Rothstein, R. A. (2002). Polish. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 686-758). Abingdon-on-Thames: Routledge.
- Scatton, E. A. (2002). Bulgarian. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 188-248). Abingdon-on-Thames: Routledge.
- Shevelov, G. Y. (2002). Ukrainian. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 947-998). Abingdon-on-Thames: Routledge.
- Short, D. (2002a). Czech. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 455-532). Abingdon-on-Thames: Routledge.
- Short, D. (2002b). Slovak. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 533-592). Abingdon-on-Thames: Routledge.
- Simpson, A. K. (2009). *The origin and development of nonconcatenative morphology* (Doctoral dissertation). University of California.

- Stephens, J. (2002). Breton. In M. Ball & J. Fife (Eds.), *The Celtic Languages* (pp. 349-409). Abingdon-on-Thames: Routledge.
- Thráinsson, H. (2002). Icelandic. In E. König & J. van der Auwera (Eds.), *The Germanic Languages* (pp. 142-189). Abingdon-on-Thames: Routledge.
- Timberlake, A. (2002). Russian. In B. Comrie & G. G. Corbett (Eds.), *The Slavonic Languages* (pp. 827-886). Abingdon-on-Thames: Routledge.
- Tuten, D. N., Pato, E., & Schwarzwald, O. R. (2016). Spanish, Astur-Leonese, Navarro-Aragonese, Judaeo-Spanish. In A. Ledgeway & M. Maiden (Eds.), *The Oxford guide to the Romance languages* (pp. 382-410). Oxford: Oxford University Press.
- Viitso, T. R. (2006). Estonian. In D. Abondolo (Ed.), *The Uralic Languages* (pp. 115-148). Abingdon-on-Thames: Routledge.
- Vincent, N. (1988). Italian. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 279-313). Abingdon-on-Thames: Routledge.
- Watkins, T. A. (2002). Welsh. In M. Ball & J. Fife (Eds.), *The Celtic Languages* (pp. 289-348). Abingdon-on-Thames: Routledge.
- Wheeler, M. W. (1988a). Catalan. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 170-208). Abingdon-on-Thames: Routledge.
- Wheeler, M. W. (1988b). Occitan. In M. Harris & N. Vincent (Eds.), *The Romance Languages* (pp. 246-278). Abingdon-on-Thames: Routledge.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, 104(19), 7780-7785.
- Zuazo, K. (2019). *Standard Basque and Its Dialects*. Abingdon-on-Thames: Routledge.

I have written this paper for the course Linguistic Universals and Diversity from the master General Linguistics at Radboud University that I followed in 2019. The original paper contained more elaborate results, which had to be left out due to length issues. If interested in these additional results, please contact the author.












Appendix

Sample of color terms and references

Table A1 contains the sample of color terms, whereas Table A2 contains the used sources. Note that the references for these sources are to be found in the reference list above.

Table A1

Sample of color terms of the fifty languages.

Language											
English	red	orange	yellow	green	blue	purple	pink	white	grey	black	brown
Frisian	read	oranje	giel	grien	blau	pears	rôze	wyt	griis	swart	brún
Dutch	rood	oranje	geel	groen	blauw	paars	roze	wit	grijs	zwart	bruin
German	rot	orange	gelb	grün	blau	violett	rosa	weiß	grau	schwarz	braun
Danish	rød	orange	gul	grøn	blå	lilla	lyserød	hvid	grå	sort	brun
Swedish	röd	orange	gul	grön	blå	lila	skär	vit	grå	svart	brun
Norwegian	rød	oransje	gul	grønn	blå	lilla	rosa	hvit	grå	svart	brun
Icelandic	rauður	appelsínugulur	gulur	grænn	blár	purpuralitur	bleikur	hvítur	grár	svartur	brúnn
Faroese	reyður	brandgulur	gulur	grønur	bláur	korkalitur	ljósareyður	hvitur	gráur	svartur	brúnur
Romansh	cotschen	oranscha	mellen	verd	blau	violet	rosa	alv	grisch	nair	brin
Ladin	cueciun	orancen	śal	vërt	brum	polpra	rosa	blanch	grisc	fosch	ros
Friulian	ros	narancin	zâl	vert	blâv	viole	rose	blanc	gris	neri	moron
Lombard	ros	aranciú	zald	verd	blö	viola	roza	bianch	gris	négher	marú
Piedmontese	ross	arancion	giàun	verd	bleu	viola	reusa	bianch	gris	nèir	maròn
French	rouge	orange	jaune	vert	bleu	violet	rose	blanc	gris	noir	marron
Occitan	roge	irange	jaune	verd	blu	violet	ròse	blanc	gris	negre	brun
Catalan	roig	taronja	groc	verd	blau	porpra	rosa	blanc	gris	negre	marró
Aragonese	royo	narancha	amariello	berde	azul	purpurencó	rosa	blanco	griso	negro	pardo
Asturian	bermeyu	naránxa	mariellu	verde	azul	moráu	rosa	blancu	gris	negru	marrón
Spanish	rojo	naránxa	amarillo	verde	azul	morado	rosa	blanco	gris	negro	marrón
Portuguese	vermelho	laranja	amarelo	verde	azul	roxo	rosa	branco	gris	negro	marron
Italian	rosso	arancione	giallo	verde	blu	viola	rosa	bianco	grigio	nero	marrone
Sicilian	russu	arancia	giarnu	viridi	bru	viola	rosa	biancu	griciu	níuru	marruni
Sardinian	ruju	aranzu	groggu	birde	blu	viola	rosa	biancu	murru	nigheddu	tabachinu
Romanian	roșu	portocaliu	galben	verde	albastru	mov	roz	alb	gri	negru	maro
Manx	jiarg	jiarg-bwee	bwee	geayne	gorrym	jiarg-gorrym	jiarg-bane	bane	lheeah	doo	dhoan
Irish	dearg	oráiste	buí	uaine	gorm	corcra	bándearg	bán	liath	dubh	donn
Gaelic	dearg	orains	buidhe	uaine	gorm	purpaidh	pinc	bàn	liath	dubh	donn
Welsh	coch	oren	melyn	gwyrd	glas	porffor	pinc	gwyn	llwyd	du	gwrn
Cornish	rudh	rudhvelyn	melyn	gwyrdh	glas	purpur	kigliw	gwynn	loos	du	gorm
Breton	ruz	orañjez	melen	gwer	glas	limestra	roz	gwenn	gris	du	melegen
Greek	kókkino	portokali	kitrino	prásino	ble	mov	roz	áspro	gkri	mávro	kafé
Albanian	kuq	portokalltë	verdhë	blerë	kaltër	purpurtë	pembë	bardhë	hirtë	zi	kaftë
Polish	czerwony	pomarańczowy	żółty	zielony	niebieski	purpurowy	różowy	biały	szary	czarny	brązowy
Czech	červená	oranžová	žlutá	zelená	modrá	fialová	růžová	bílá	šedá	černá	hnědá
Slovak	červená	oranžová	žltá	zelená	modrá	fialová	ružová	biela	šedá	čierna	hnedá
Slovene	rdeča	oranžna	rumena	zelena	modra	vijolična	roza	bela	siva	črna	rjava
Serbian	crven	narandžast	žut	zelen	modar	ljubičast	ružičast	beo	siv	crn	braon
Bulgarian	červén	oránžev	žlt	zelen	sin	violétov	rózov	bjal	siv	čeren	kafjáv
Macedonian	crven	portokalov	žolt	zelen	sin	purpuren	rozov	bel	siv	crn	kafen
Russian	krásnyj	oránževyj	žóltij	zeljónyj	sinij	púrpurnyj	rózovyj	bélyj	séryj	čórnyj	koríčnevij
Ukrainian	červónyj	oránževyj	žóvtij	zelényi	sýnij	púrpurovyj	rožévij	bílyj	síryj	čórnyj	korýčnevij
Belarusian	čyrvóny	aránžavy	žóuty	zjaljóny	síni	purpúrny	ružóvy	bjély	šéry	čórny	karýčnjevy
Lithuanian	raudona	oranžinė	geltona	žalia	mėlyna	violetinė	rožinė	balta	pilka	juoda	ruda
Latvian	sarkans	oranžs	dzeltens	zaļš	zils	purpursarkans	rozā	balts	pelēks	melns	brūns
Estonian	punane	oranž	kollane	roheline	sinine	purpurne	roosa	valge	hall	must	pruun
Finnish	punainen	oranssi	keltainen	vihreä	sininen	purppura	pinkki	valkea	harmaa	musta	ruskea

Hungarian	piros	narancsszín	sárga	zöld	kék	bíbor	rózsaszín	fehér	szürke	fekete	barna
Basque	gorri	laranja	hori	berde	urdin	more	arrosa	zuri	gris	beltz	marroi
Maltese	ahmar	orangjo	isfar	ahdar	ikhal	vjola	roża	abjad	griz	iswed	kannella

Table A2

References used for each of the fifty languages.

Language	Reference	Language	Reference
English	König, 2002	Manx	Broderick, 2002
Frisian	Hoekstra & Tiersma, 2002	Irish	Mac Eoin, 2002
Dutch	de Schutter, 2002	Gaelic	Gillies, 2002
German	Eisenberg, 2002	Welsh	Watkins, 2002
Danish	Haberland, 2002	Cornish	George, 2002
Swedish	Andersson, 2002	Breton	Stephens, 2002
Norwegian	Askedal, 2002	Greek	Hoenigswald, 2015
Icelandic	Thráinsson, 2002	Albanian	Demiraj, 2015
Faroese	Barnes & Weyhe, 2002	Polish	Rothstein, 2002
Romansh		Czech	Short, 2002a
Ladin	Haiman, 1988	Slovak	Short, 2002b
Friulian		Slovene	Priestly, 2002
Lombard		Serbian	Brown, 2002
Piedmontese	Benincà, Parry, & Pescarini, 2016	Bulgarian	Scatton, 2002
French	Harris, 1988	Macedonian	Friedman, 2002
Occitan	Wheeler, 1988b	Russian	Timberlake, 2002
Catalan	Wheeler, 1988a	Ukrainian	Shevelov, 2002
Aragonese		Belarusian	Mayo, 2002
Asturian	Tuten, Pato, & Schwarzwald, 2016	Lithuanian	Balode & Holvoet, 2001b
Spanish		Latvian	Balode & Holvoet, 2001a
Portuguese	Parkinson, 1988	Estonian	Viitso, 2006
Italian	Vincent, 1988	Finnish	Abondolo, 2006a
Sicilian	Ledgeway, 2016	Hungarian	Abondolo, 2006b
Sardinian	Jones, 1988	Basque	Zuazo, 2019
Romanian	Mallinson, 1988	Maltese	Kaye & Rosenhouse, 1997