Synonymy in the Language of Colour

Dimitris Mylonas [1,*], Alexandros Koliousis [1] and Mari Uusküla [2]

[1] New College of the Humanities, Northeastern University, UK; dimitris.mylonas@nchlondon.ac.uk, alexandros.koliousis@nchlondon.ac.uk
[2] School of Humanities, Tallinn University, Estonia; muuskyla@tlu.ee
* Corresponding author: dimitris.mylonas@nchlondon.ac.uk

Abstract

We explore synonyms in colour naming within and across three languages, British English, Estonian and Greek, using data collected from a crowdsourcing experiment. We identified 30 common lexical colour categories in British English, 41 in Estonian and 29 in Greek, where no one category was fully contained within others. The synonymy analysis within languages revealed that the highest degree of overlapness was found for a pair of dark reddish loanwords in English (maroon and burgundy) and in Greek (*bissini* and *bornto*) that were absent in Estonian. The synonymy of two purplish categories in Estonian (*lilla* and *tumelilla*) and Greek (*mov* and *lila*) was also prominent but in English purple and lilac were more separated. The investigation of synonymy across languages revealed similar graph properties for all pairs of languages (British English – Estonian; Estonian – Greek; Greek – British English). Our results suggest that the degree of synonymy in the language of colour is influenced by cross-cultural transfer of loan words.

Keywords: *colour, naming, synonyms, loanwords, cross-cultural*

INTRODUCTION

In natural language, a colour can often be named by two or more synonymous words. Synonyms that share exactly the same colour category are exceptionally rare but near synonyms, with an extent of overlapping colours, are numerous. From an onomasiological perspective, when a large, well-established colour category (e.g., red), contains a smaller colour category (e.g., crimson), the two colour names are considered synonyms. But from a semasiological perspective, their relation would be that of hyponymy, because the crimson is a sub-category or hyponym of red, whereas red contains crimson hues and is therefore a hypernym of crimson (MacLaury, 1997). Therefore, the standard synonymy test of substitution, where one word can be replaced by another word without changing its meaning, is not appropriate for colour naming: we can say that crimson is a kind of red, but we cannot say that red is a kind of crimson

The problem of finding synonyms is even more complicated when we consider languages with a different number of colour categories and their corresponding basic colour terms. For example, Greek and Italian both contain at least 12 basic colour terms (Athanasopoulus, 2009; Uusküla, 2014). But in English, as a counterexample, Mylonas and MacDonald (2016) suggested the augmentation of the English inventory from the 11 basic colour terms (Sturges & Whitfield, 1995) to 13 terms, adding lilac and turquoise. The underlying processes for developing colour naming systems remain unsettled, with no consensus as to whether lexical colour categories are formed under the influence of perceptual mechanisms (Berlin & Kay, 1969/1991) or cultural communication needs (Davidoff, 2015).

In this study, we focus on quantifying the degree of near synonymy between colour names from a semasiological perspective for a set of three colour languages. Our synonymy methodology comprises two steps: (i) map common colour names to distinguishable colour categories, where no

one category is fully contained within others; and (ii) identify overlapping categories in colour space as synonyms in name space, thus quantifying the degree of near synonymy.

METHODS

Data Collection

A crowdsourcing colour naming experiment (https://colournaming.com) was designed to collect unconstrained names for 600 in total samples from the Munsell Renotation Data set, including eleven achromatic samples (Mylonas & MacDonald, 2010). Participation was voluntary and anonymous. Colour stimuli subtending a visual angle of about 3 degrees at a viewing distance of about 50cm were presented sequentially against a mid-neutral background with a black outline of 1 pixel in random order. Typed responses along with the typing onset delay were recorded. Each participant was free to use any colour descriptor, either a single word, or a compound, or term(s) with modifiers or qualifiers. All participants were screened for possible colour vision deficiencies with a web-based Dynamic Colour Vision Test (Barbur, 2004).

In this study, we consider 10,000 raw responses from 500 British English (EN), 10,000 responses from 333 Estonian (EE) and 10,000 responses for 532 Greek (GR) participants. We exclude disruptive observers offering incomplete, numerical and responses written in languages other than the language of the instructions, and observers with possible colour deficiencies (EN: 9.6%, EE: 2.8%, GR: 18.7%). Typographic conventions were replaced with spaces, leading and trailing spaces were removed, and all multi-character spaces were reduced to single spaces. Capitalisation was ignored. Common spelling errors (e.g., 'fusia' instead of 'fuchsia') were corrected with supervision. To measure the synonymy of colour terms we restrict our analysis to single word colour names which were produced at least 10 times in our data to give us confidence in their distribution. This filtering resulted in a dataset comprising 443 English, 276 Estonian and 343 Greek speaking respondents with a mean age of 32 (SD=12), 41 (SD=11), 31 (SD=9) years, respectively.

English and Greek are both Indo-European languages, belonging to different branches of that language family. However, Estonian is a Finno-Ugric language and its word-formation rules are different from English and Greek. For example, English and Greek both contain more object-derived colour names (Athanasopoulos, 2009; Mylonas & MacDonald, 2016) and therefore translating between English and Estonian can sometimes cause confusion (Uusküla, 2019). There are also orthographic differences: in English and Greek modifiers are separated by a space but in Estonian, modifiers are used as compound names. Due to rich word-formation in Estonian, Estonian subjects tended to offer more colour terms per subject. This is a common feature of all Finno-Ugric languages in which colour naming has been studied (Uusküla et al., 2012).

Data Modelling

We establish common colour names in each language that are reliably distinguishable from other names in colour space using a probabilistic model based on Maximum a Posteriori (MAP) (Mylonas & MacDonald, 2016). More than one colour names may be offered for a colour, but the MAP model favours a colour name with the highest frequency among observers to maintain congruence between observed and predicted data. So, colour names that correspond to larger and more consistent colour categories tend to subsume smaller and inconsistent sub-categories.

To identify synonyms for colour names, the first relationship we explore is the conditional probability $P(C|n)$ that describes the likelihood of a given colour stimulus C, being referred to by each

distinguishable colour name, *n*. We can then express the degree of near synonymy between colour names as the amount of overlap between pairs of probability distributions using the Hellinger distance (H). Hellinger distance is symmetric and obeys the triangle inequality.

The metric space defined by Hellinger distance borrows itself to a network view. We can view the colour naming space of a given language as an undirected weighted graph where nodes are centroids of colour names in CIELAB; and the weighted edge between two colour names is their Hellinger distance. Our network view enables us to analyse the colour naming space both within and across languages. Within a language, we can use measures of centrality (e.g., degree and closeness) to rank nodes and identify key infrastructure nodes. For example, closeness centrality (Bavelas, 1950) allows us to identify a subset of colours that are closer to all other colours. Across languages, we can use graph edit distance measures to identify how similar two graphs are. In this work, we focus primarily on node edit distance, defined as the distance between two synonyms across languages in CIELAB

RESULTS AND DISCUSSION

Synonymy within languages

Using the MAP procedure, we identify 30 distinguishable lexical colour categories in English, 41 in Estonian and 29 in Greek. Our results agree with previous estimates on the number of 30-50 distinct colour names, as identified by native speakers, that fit within colour space (Chapanis, 1965; Derefeldt & Swartling, 1995; Griffin & Mylonas, 2020). The larger number of Estonian common colour names reflects that Estonian speakers use modifiers as compounds in single word forms. Figure 1 shows these common colour terms in English, Estonian and Greek as graph networks. Synonyms with each language are connected with lines and the width of each line corresponds to their degree of synonymy. In British English, the strongest degree of synonymy was found between maroon and burgundy (Hellinger distance, H=0.6), followed by peach and salmon (H=0.5) and cyan and turquoise (H=0.5). Both maroon (*marron*) and burgundy are loan words from French and Latin describing a dark reddish colour region. French is also the origin of peach (*pêche*) and salmon (*saumon*) referring to pale orange-pinkish colours. Cyan (*kyano*), originating from Greek, and turquoise (turquois), originating from French meaning Turkish, both refer to greenish-blue colours. The degree of synonymy between lilac and purple is considerable smaller in British English than in Estonian and Greek supporting the candidacy of lilac as a basic colour term (Mylonas & MacDonald, 2016).

In Estonian, *hallikassinine* (bluish grey) and *sinakashall* (greyish blue) terms were the strongest synonyms (H=0.6), followed by *sinepikollane* (mustard yellow) and *rohekaskollane* (greenish yellow, H=0.5) and *lilla* (purple) and *tumelilla* (dark purple, H=0.5). The synonymity between *hallikassinine* and *sinakashall* is predictable given that are compounds of the same words in different order. However, given that the two colour terms refer to different areas in colour space – one to grey and the other one to blue – these are not full synonyms. *Sinepikollane* and *rohekaskollane* both refer to the same base colour: yellow. However, *sinepikollane* has a slightly brownish overtone and *rohekaskollane* refer to a greenish yellow colour. *Lilla* and *tumelilla* both belong to the same colour category: purple. *Tumelilla* (dark purple) is a specification of *lilla* (purple), because you can use the modifier *tume* (dark) to specify the type of purple.

Similar to English, in Greek, the closest synonyms were a pair of dark reddish colours named as *bissini* and *bornto* (H=0.7), followed by two whitish terms *aspro* and *lefko* (H=0.6) and two purplish terms *lila* and *mov* (H=0.5). *Bissini* (*βύσσος*) comes from ancient Greek while *bornto* (*bordeaux*) is a loan word for wine red from French. *Lefko* is an ancient Greek word that is used more often to describe whitish objects with some specular component (e.g., hair) while *aspro* is a more general term that comes from Latin. *Mov* (mauve) comes from French while *lila* from Turkish.

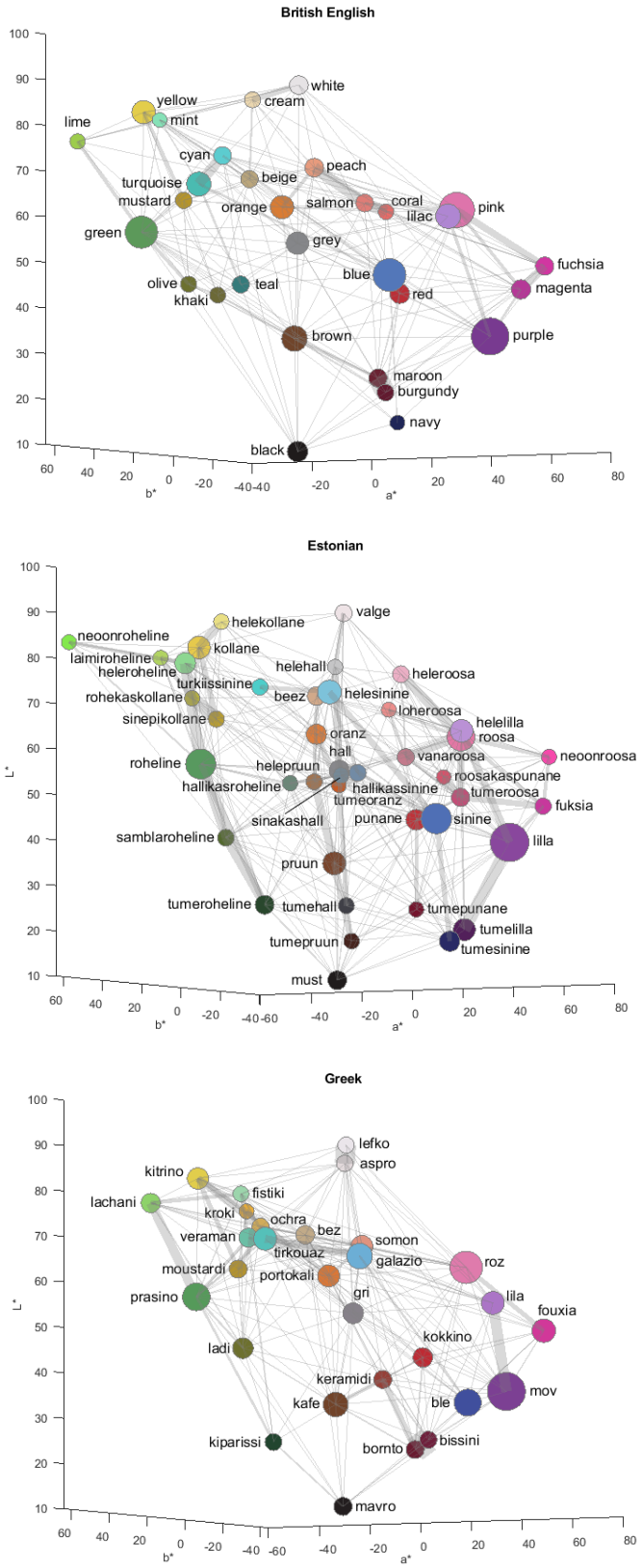

Figure 1: Common colour names in British English (top, *n*=30), Estonian (middle, *n*=41) and Greek (bottom, *n*=29) in CIELAB. The colour and location of the discs corresponds to the coordinates of the centroids of the colour categories. The size of the discs corresponds to their frequency in the online experiment. The grey lines between the discs link colour categories that share common colour samples. The width of the lines indicates the degree of overlapness between pairs of colour categories.

In terms of closeness centrality within languages, greyish categories were ranked first in British English (grey; 0.88) and in Estonian (*hall*; 0.83) while in Greek was ranked 3rd (*gri*, 0.82) because of their central position in colour space, they were closer to all other nodes. The greenish categories were also found at the top of the ranks (green-EN: 0.83; *roheline*-EE: 0.75; *prasino*-GR: 0.78): because of their relative larger volume in colour space. An interesting finding was that beige-ish categories were found at the top of the ranks too (*bez*-GR: 0.91; beige-EN:0.72; *beez*-EE: 0.68) due to the large number of colour names (e.g., peach, salmon, cream) offered to name this region.

Synonymy across languages

Using the same procedure of measuring the overlapness between pairs of categories we were able to determine the similarity between pairs of graphs of each language (British English – Estonian; Estonian – Greek; Greek – British English). The average degree of synonymy between British English and Estonian was H=0.21 (SD=0.19), while against Greek was H=0.19 (SD=0.2). The mean synonymy between Estonian and Greek was H=0.19 (SD=0.16).

In terms of node edit distance defined as the mean CIEDE 2000 colour difference between synonyms across languages, the British English language of colour is closer to Greek (DE00=5.01, SD=5.25) than to Estonian (DE00=5.65, SD=4.81) while the Estonian is closer to Greek than to British English (DE00=5.23, SD=3.45). Yet the differences are small and not significant indicating that our three languages have similar graph properties.

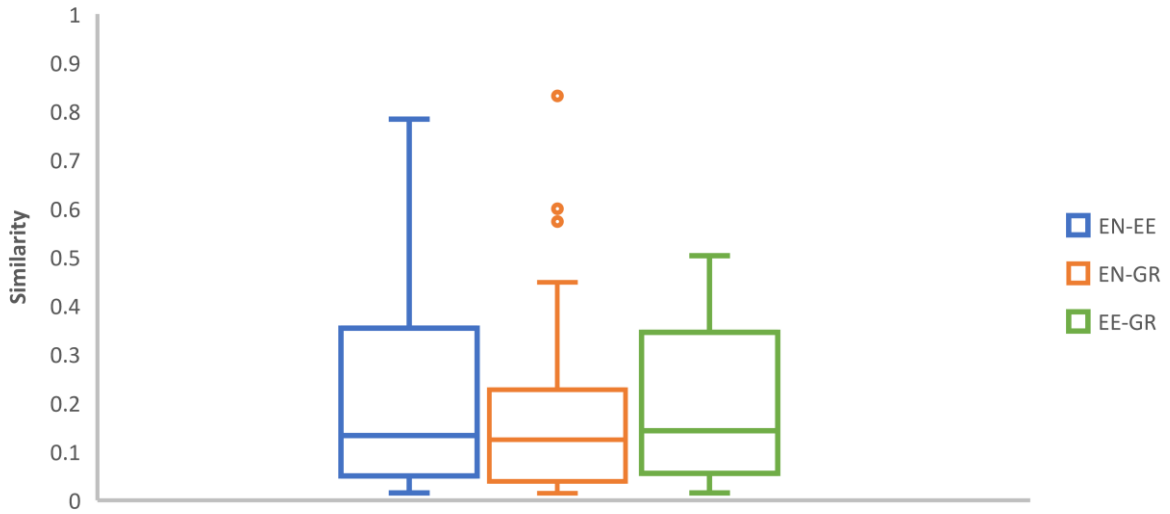

Figure 2: Degree of synonymy across pairs of languages, left: British English-Estonian, middle: British English-Greek, right: Estonian-Greek.

CONCLUSIONS

In summary, the referential meaning of colour names can be mapped on colour space where we can determine the degree of synonymy between colour names based on the overlapness of their corresponding categories. Our investigation of colour synonymy using colour naming data in British English, Estonian and Greek from a crowdsourcing experiment revealed that synonymy within a language of colour is stronger when loanwords for object colours are introduced in their colour lexicons. In contrast, the analysis of synonymy across languages showed consistent variability with similar graph properties for all pairs of languages. On the whole, our findings suggest that synonymy in the language of colour is modulated by cross-cultural transfer of object colours.

ACKNOWLEDGEMENTS

We thank all volunteers who have contributed to this study as participants in the online experiment. Dimitis Mylonas and Alexandros Koliousis were supported by a TIER 1 Seed Grant from Northeastern University. Mari Uusküla is thankful to Tuuli Järmut for help in data collection.

REFERENCES

Athanasopoulos, Panos. 2009. Cognitive Representation of Colour in Bilinguals: The Case of Greek Blues. *Bilingualism: Language and Cognition* 12 (01): 83–95. https://doi.org/10.1017/S136672890800388X.

Barbur, J. L. 2004. Double-Blindsight Revealed through the Processing of Color and Luminance Contrast Defined Motion Signals. *Progress in Brain Research* 144: 243–59. https://doi.org/10.1016/S0079-6123(03)14417-2.

Bavelas, Alex. 1950. Communication Patterns in Task-Oriented Groups. *The Journal of the Acoustical Society of America* 22 (6): 725–30. https://doi.org/10.1121/1.1906679.

Berlin, B., and P. Kay. 1969/1991. *Basic Color Terms: Their Universality and Evolution*. The David Hume Series. Stanford, Calif.: Center for the Study of Language and Information.

Chapanis, A. 1965. Color Names for Color Space. *American Scientist* 53: 327–46.

Davidoff, J. 2015. Color Categorization across Cultures. In: *Handbook of Color Psychology*, ed. A.J. Elliott, Mark D. Fairchild, and A. Franklin, 259–78. Cambridge: Cambridge University Press.

MacLaury, R. E. 1997. *Color and Cognition in Mesoamerica: Constructing Categories as Vantages*. Facsimile edition. Austin: University of Texas Press.

Mylonas, D., and L. MacDonald. 2010. Online Colour Naming Experiment Using Munsell Samples. In: *5th European Conference on Colour in Graphics, Imaging, and Vision and 12th International Symposium on Multispectral Colour Science*, 27–32. Joensuu, Finland: IS&T.

Mylonas, D., and L. MacDonald. 2016. Augmenting basic colour terms in English. *Color Research & Application*, *41*(1), 32–42. https://doi.org/10.1002/col.21944.

Derefeldt, G., and T. Swartling. 1995. *Colour Concept Retrieval by Free Colour Naming. Identification of up to 30 Colours without Training.* Displays 16 (2): 69–77.

Griffin, L. D., and D. Mylonas. 2019. *Categorical Colour Geometry*. PLOS ONE 14 (5): e0216296. https://doi.org/10.1371/journal.pone.0216296.

Uusküla, M. 2014. Linguistic Categorisation of BLUE in Standard Italian. In: *Colour Studies: A Broad Spectrum*, ed. W. Anderson, C. P. Biggam, C. Hough, C. Kay. Amsterdam/Philadelphia: John Benjamins, 67–78.

Uusküla, M. 2019. Translation of colour terms: An empirical approach toward word-translation from English into Estonian. *ESUKA - JEFUL*, 10-2, 69–84.

Uusküla, M., L. Hollman, and U. Sutrop 2012. Basic colour terms in five Finno-Ugric language and Estonian Sign: A comparative study. ESUKA - JEFUL, 3-1, 47–86.

Sturges, J., and A. Whitfield. 1995. Locating Basic Colours in the Munsell Space. *Color Research & Application,* 20 (6): 364–76.