Colournamer – a synthetic observer for colour communication

Dimitris MYLONAS[1], Jonathan STUTTERS[1], Valero DOVAL[2], Lindsay MACDONALD[3]
[1] Wellcome Laboratory of Neurobiology, University College London
[2] El Atelier Company, Valencia, Spain
[3] Dept. of Medical Physics, University College London

ABSTRACT

Colour specification is not only the domain of technologists but is also an important process for anyone who needs to communicate about colour in the multilingual Internet environment. We have developed an online application *Colournamer,* a synthetic observer 'trained' by the participants' responses, to facilitate colour communication between different cultures. At present it supports English, Greek, Spanish and German.

1. INTRODUCTION

Various methodologies for developing colour-naming models from experimental data have been proposed. In most cases, however, these models are constrained to a small number of colour names and assume a 'universal' colour categorisation (Benavente *et al*, 2008; Lin *et al*, 2001). In such models the use of each colour name is restricted to a unique colour category, which means that when colour names are translated into other languages the colour categories remain the same. This has the advantage that it requires partitioning the colour space only once and then simply translating the selected words to each language. Universal colour naming models must assume that the chosen colour name represents the same colour category on a global scale, with a firm commitment by all the involved cultures. This assumption is inaccurate, and results in a colour space that is only partially mapped by colour language (Berlin and Kay, 1991; Gage, 1993).

Our *Colournamer* application adopts an alternative methodology based on an online colour-naming model that is distributed worldwide (Fig. 1). It is composed of multiple 'culture dependent' lexical sub-models, each of which is based on the same numerical 'culture independent' colour model (Mylonas *et al*, 2010). This global but relativistic framework, where each colour name is bound to a colour category in a particular cultural context, supports more subtle colour identification. This is in accord with recent scientific findings that show the influence of language on categorical perception. What we have developed, in effect, is a synthetic observer that is able to predict a colour name with the highest probability of agreement with the judgements of thousands of participants in our ongoing colour naming experiment. This enables the system to return the most likely colour sample when a user enters a colour name, and conversely also to respond with the most likely name in a given language when a user chooses a colour sample.

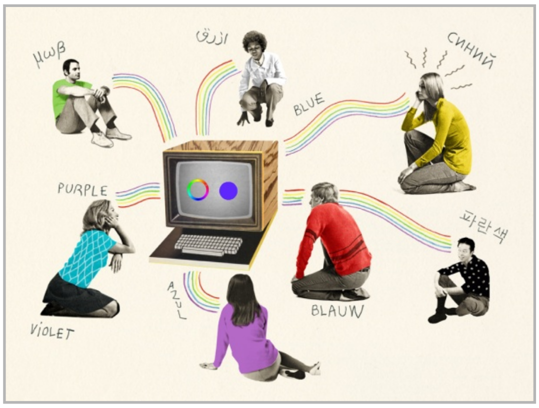

Figure 1: Illustration of Colournamer framework, by Valero Doval.

2. EXPERIMENTAL METHOD

A custom online colour naming experiment at http://colournaming.com was designed to collect broad sets of multi-lingual colour names with their corresponding colour ranges in sRGB and

Munsell specifications. The 600 test stimuli were selected from the Munsell Renotation Data, and colours were specified in the sRGB colour space. Samples outside the sRGB gamut were excluded. In each experimental session an observer names 20 samples, selected at random from the 600 test colours (Mylonas & MacDonald 2010).

Over the past four years (2009–13), the server has gathered responses from over 3,000 participants. This has produced a dataset of over 60,000 colour names in eleven languages: English, Greek, Spanish, German, Catalan, Italian, simplified Chinese, traditional Chinese, Korean, French and Danish. Associated metadata include the language, cultural background and possible colour deficiency of each observer, and the hardware/software configuration and viewing conditions prevailing in the experimental session. To validate our methodology, we compared the centroids of the 11 English basic colour terms (BCTs) with previous studies conducted in controlled laboratory settings and the 27 most frequent chromatic colour words against the results of the web-based experiment of Moroney (2003). Both tests have produced satisfactory agreement (Mylonas & MacDonald 2010). A colour-naming model was developed, with maximum *a posteriori* (MAP) parameter estimation (Mylonas *et al*, 2010).

3. DEVELOPMENT OF COLOURNAMER

The data for the four most popular languages of the experiment – English, Greek, Spanish and German – were analysed after removing the responses of observers with possible colour deficiency. From 5428 total responses in English we identified 1166 unique colour names, and from 4274 responses in Greek we identified 984 unique colour names. For Spanish and German the total responses were 2960 with 928 unique colour names and 4280 with 1161 unique colour names respectively. Using our MAP algorithm we classified the colour coordinates of all test stimuli to determine which colour names would retain their identity across languages. This could apply when a user chooses a colour name to return the most likely sample, or conversely when a user selects a colour sample to return the same colour name. We identified 30 predominant colour names in English, 27 in Greek, 24 in Spanish and 29 in German, as shown in Fig. 3.

Colournamer is a synthetic observer using participants' responses and a 'democratic' probabilistic algorithm to assign a colour name automatically. Through a simple graphic user interface (Fig. 2) it enables a user to enter a name to return the most likely colour sample, and conversely to choose a colour sample to respond with a word cloud of the most likely names. The font size of each colour name is related to its probability, while the colour names are plotted using their coordinates in the a^*-b^* plane. The web app was developed with Python and the visualisation with canvas of HTML-5. Recently, we have used the new web audio APIs of HTML-5, so visitors with an up-to-date browser will be able to listen to a recorded voice naming the selected colour samples. Users are also invited to validate the predictions of the application to improve its performance.

Figure 2: Screenshot of the Colournamer application showing the most probable English names corresponding to a selected colour.

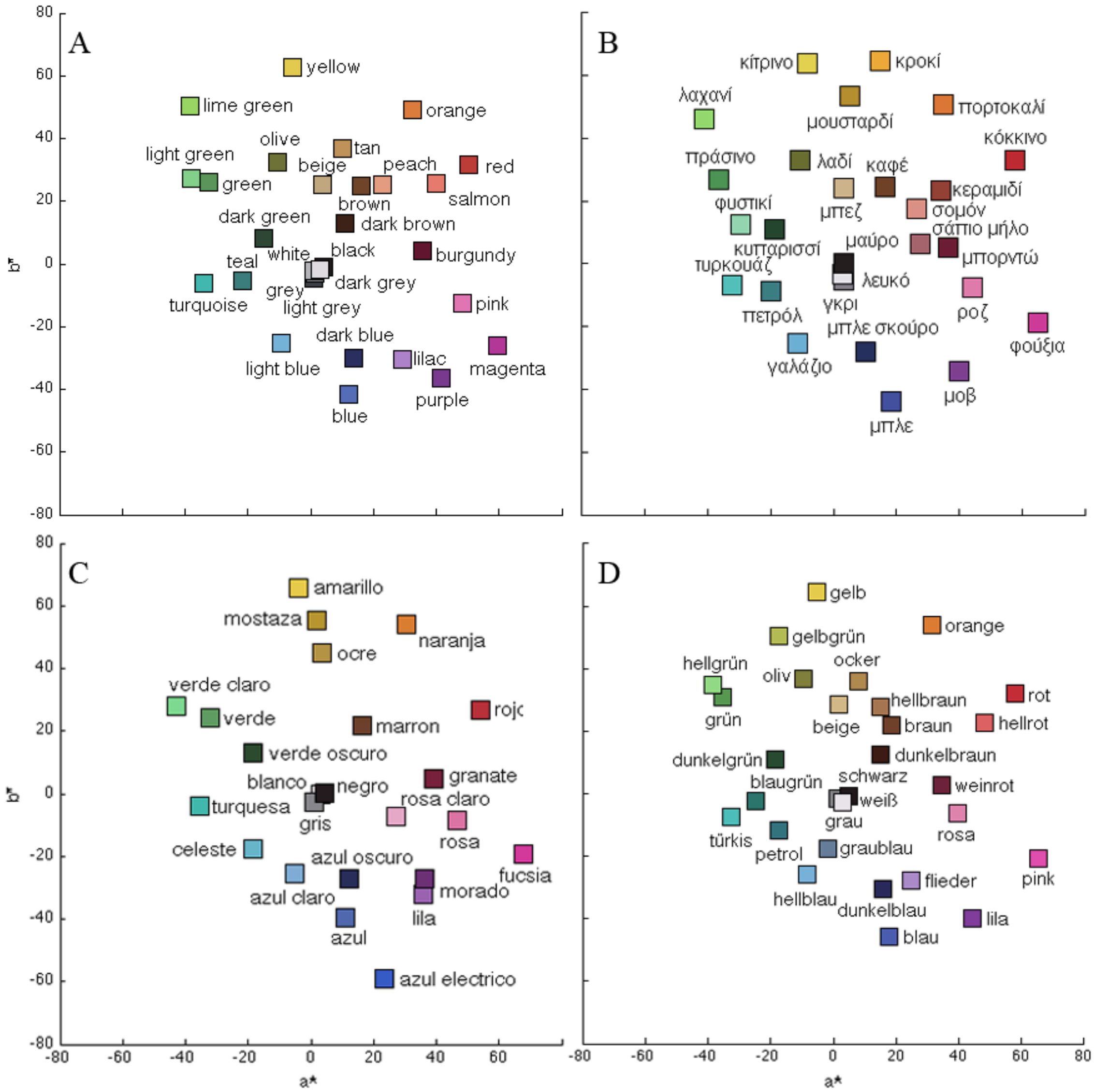

Figure 3: Centroids of predominant colour names in the CIELAB a-b* plane for: (A) English; (B) Greek; (C) Spanish; and (D) German datasets.*

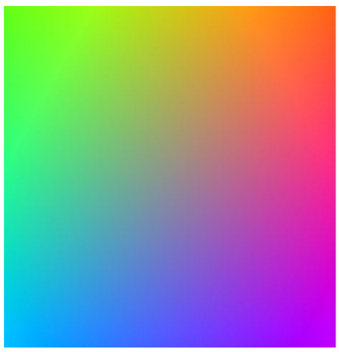

Figure 4: Synthetic image Van de Weijer et al, 2009

Using the same probabilistic model and the same four training sets we classified the 320 samples of Munsell's hue-value surface and segmented a synthetic image (Fig. 4). In the results in Fig. 5 we can see interesting similarities and differences. The synthetic image was segmented into 19 English, 18 Greek, 16 Spanish and 20 German colour categories. English and Greek observers used the colour names 'lime green' and 'lahani' (λαχανί) to describe a region between green, turquoise and yellow, while for the same area Spanish and Germans used the modifiers of green 'verde claro' and 'hellgrun' (light green). Greek observers used the basic term (Androulaki *et al*., 2006) 'galazio' (γαλάζιο) to describe a large area between blue and turquoise which was not shared with the other tested languages. In German an equivalent area was described as 'hellblau' (light blue) with a hue angle different from 'blau' (blue). This is interesting because we usually associate 'light' as a modifier of lightness only. The same trend was observed for English too, but covering a smaller area. A large area between blue and green was named as ‘turquoise’ in English, ‘τυρκουάζ’ in Greek, ‘turquesa’ in Spanish and ‘türkis’ in German, and it appears to be a strong candidate for status as a twelfth basic colour term.

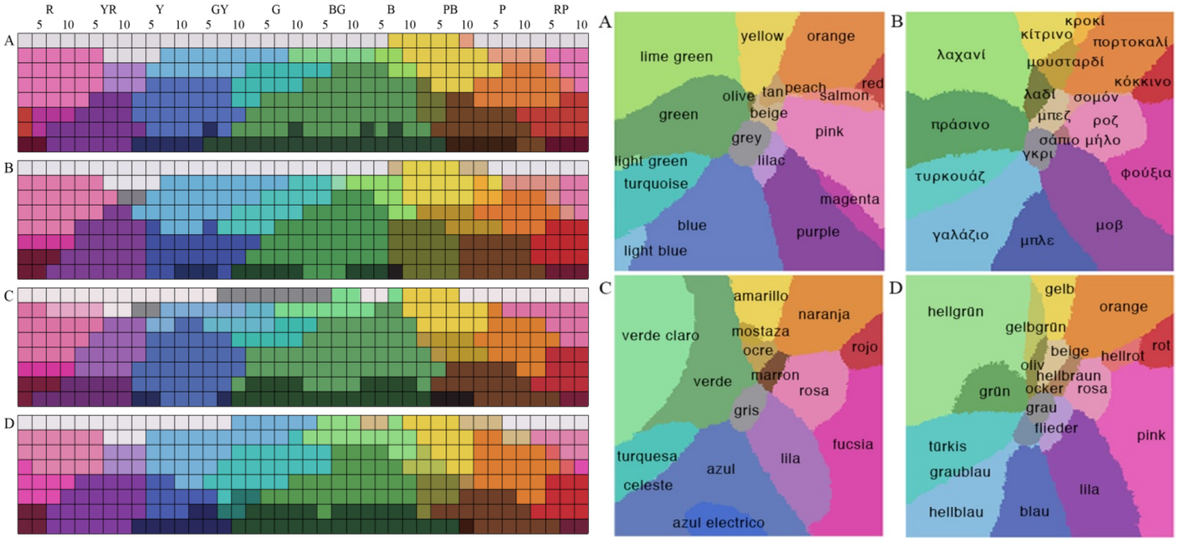

Figure 5: Classification of (left) Munsell 320-colour array of value vs hue and (right) synthetic image by observer-based colour names in: (A) English; (B) Greek; (C) Spanish; (D) German. Each category is represented by the colour of its centroid (see Fig. 3).

In conclusion, we have demonstrated that a flexible colour naming architecture adapted to the communication needs in each language/culture has significant advantages over universal models, because it is able to represent more consistently the native colour concepts of its users. It offers a route to culture-dependent interpretation of colour names.

REFERENCES

Androulaki, A., N. Gomez-Pestana, C. Mitsakis, J. Jover, K. Coventry, and I. Davies. 2006. Basic colour terms in Modern Greek: Twelve terms including two blues. *J. Greek Linguistics*, 7(1):3–47.

Benavente, R., M. Vanrell and R. Baldrich. 2008. Parametric fuzzy sets for automatic color naming. *JOSA A*, 25(10):2582-2593.

Berlin, B. and P. Kay. 1991. *Basic Color Terms: Their Universality and Evolution*. University of California Press.

Gage, J. 1993. *Color and Culture: Practice and Meaning from Antiquity to Abstraction*, Boston MA: Little, Brown and Company.

Lin, H., M.R. Luo, L.W. MacDonald and A.W.S. Tarrant. 2001. A cross-cultural colour-naming study. Part III – A colour-naming model. *Color Res. & Appl.*, 26(4):270-277.

Moroney, N. 2003. Unconstrained web-based color naming experiment. Color Imaging VIII: Proc. Hardcopy and Applications, *Proc. SPIE,* 5008:36-46.

Mylonas, D. and L.W. MacDonald. 2010. Online Colour Naming Experiment using Munsell Colour Samples. *Proc. 4th Colour in Graphics, Imaging & Vision*, 27-32.

Mylonas, D., L.W. MacDonald and S. Wuerger. 2010. Towards an online colour naming model, *Proc. 18th IS&T/SID Color Imaging Conf.*, 140-144.

Van de Weijer J., C. Schmid, J. Verbeek and D. Larius. 2009. Learning Color Names for Real-World Applications. *IEEE Trans. on Image Processing*, 18(7):1512-1523.

Address: Dimitris Mylonas, Wellcome Laboratory of Neurobiology,
University College London, Gower Street, WC1E 6BT
Email:d.mylonas@ucl.ac.uk