Spectrographic analysis of colourants of cultural items: from a qualitative to a semi-quantitative data treatment through BCTs

Guido Frison [1,*], Maurizio Aceto [2], Angelo Agostino [3], Dimitris Mylonas [4] and Alberto Calatroni [5]

[1] UCL Anthropology Department, 14 Taviton Street, London WC1H0BW, UK; frison.guido@gmail.com
[2] Dipartimento di Scienze e Innovazione Tecnologica, Università degli Studi del Piemonte Orientale, 15121 Alessandria, Italy; maurizio.aceto@uniupo.it
[3] Dipartimento di Chimica, Università degli Studi di Torino, 10125 Torino, Italy; angelo.agostino@unito.it
[4] New College of the Humanities at Northeastern. Assist. Professor in Computer Science; dimitris.mylonas@nchlondon.ac.uk
[5] Lucerne University, Senior scientist; calatroni.alberto@gmail.com
* Corresponding author: frison.guido@gmail.com

Abstract
We propose to use current spectrographic techniques for detecting colourants in cultural items in a semi-quantitative way, by systematically utilising the notions of colour term, basic colour term (BCT), and by sampling the coloured areas of each BCT. Given for example two manuscripts (MSS), criteria for sampling the areas of the illuminations of a given BCT, and the detected colourants of those areas, it is possible to estimate a degree of colourant difference. Further, a unique identifier of the item colourants is proposed. For achieving this aim we discuss some steps of the analytical procedure, and borrow from ecology three homologous concepts for detecting colourants, i.e.: for a given BCT, the colourant richness, the colourant abundance distribution, and the evenness of this distribution. Further, we explore a procedure, the input of which is a generic picture, and the output is a map of the same object giving the BCTs distribution.

***Keywords**: colour term, BCT, colourant richness, colourant abundance distribution, colourant evenness*

INTRODUCTION

Spectrographic techniques for detecting colourants and other materials used in visual arts are based on three conceptual notions, namely, colour terms of the selected stained spots, spectrographic data, and colourants. However, colour terms have never been systematically utilised. Moreover, spectrographic techniques have been used since their beginnings in the early 1990s in a purely qualitative way, without a sample criterion concerning the possible presence of rarely used colourants.

We propose to use current spectrographic procedures in a semi-quantitative way by systematically utilizing the notions of 'colour term' and 'basic colour term' (BCT). Given, for example, two MSS, and criteria for sampling the areas of the illuminations of a given BCT, it is possible to estimate a quantitative degree of their difference. Further, a unique identifier of the item colourants is proposed.

For achieving this aim we will provide a brief introduction to the following empirical concepts of a generic cultural item, namely, the colourant richness, or number of colourants used for each BCT, its colourant abundance distribution (CAD), and the evenness of this distribution, or the degree of variability in colourant abundances. These notions are already used in other disciplines, in particular in ecological studies on biological diversity.

One of us suggested a systematically utilization of the notions of colour term and its corresponding BCT[1] for each analysed stained spot. The modern English BCTs are the following in alphabetical order

[1]According to Berlin and Kay (1969: 5-7), a BCT should have the following four properties:
i) "It [the BCT] is monolexemic; that is, its meaning is not predictable from the meaning of its parts";
ii) "Its signification is not included in that of any other color term";
iii) "Its application must not be restricted to a narrow class of objects";

only for convenience: BLACK, BLUE, BROWN, GREEN, GREY, ORANGE, PINK, PURPLE, RED, WHITE, YELLOW (in capital letters to differentiate them from colour terms). The difference between colour term and BCT is paramount: for example, 'red' is a BCT, but not 'red-orange' or 'dark red' or 'light red', that are included in the RED BCT.

The treatment of colour terms and BCTs triggered a further working hypothesis of creating an application capable of capturing an image of an illumination and producing the distribution of the BCTs of the miniature itself as an output.

The article is divided in two sections: the first deals briefly with the notions listed above, and the second one examines only few important problems, that we encountered in the course of the analytical process. Further, we illustrate the design of a procedure, the input of which is a generic picture of a given illumination of a MS or a picture, and the output is a picture of the same object giving the BCTs distribution.

A SHORT OVERVIEW OF THE NOTIONS OF COLOURANT ABUNDANCE DISTRIBUTION (CAD), RICHNESS AND EVENNESS

The detection of the colourant abundance distribution of a BCT is one thing, and providing a synoptic measure of the diversity of colourants of the same BCT of two items represents a second different challenge. A further problem is detectability: not all colourants are equally easy to detect because any analytical technique has definite and specific limits.

Given a cultural item and one of the 11 BCTs, the CAD refers to a vector of colourants, and their absolute and/or relative abundances (see below Table 1). The 11 CADs, one for each BCT, represent a simple but complete information which can be called *colourant identifier of the cultural item*.

Relat. freq.	Azurite	Indigo	Ultram. blue	Madder/ Indigo	Indigo/ Yell. Col.	Indigo/ Ultram.	Unk-nown	**Σ relat & abs. freq.**	Statistical units (Number of MSS)
a_i	7.14	35.71	38.1	14.29	4.76	0	0	**(100) 42**	3 Late antiquity MSS
b_i	0	20.69	74.14	0	0	3.45	1.72	**(100) 58**	7 Medieval MSS

Table 1: BLUE BCT: two colourant abundance distributions from 10 MSS of the Paris Bibliothèque Nationale de France. The CADs of two Syriac collections of MSS (relative (a_i and b_i) and absolute colourant abundances).

We wish to underline that only the relative abundances are conserved, whether the sample sizes are appropriate. N.B.: we arbitrarily consider as a colourant not only the pure colourant (e.g. azurite, indigo, etc.), but also their mixtures, for increasing the number of 'colourants'. Many different graphical procedures are used to display the equivalent of the CADs, see McGill (2011).

iv) "It must be psychologically salient for informants".

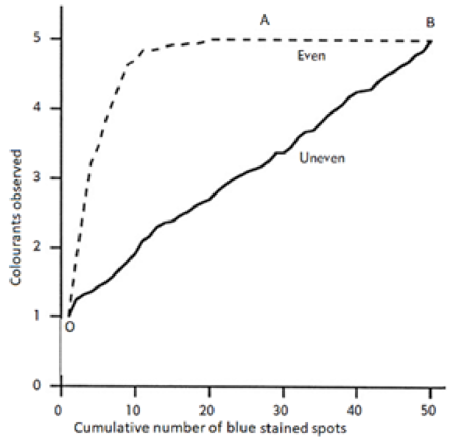

Figure 1: The observed colourant richness of a cultural item in function of the cumulative number of the (BLUE) stained spots. Courtesy by Anne Magurran (2004: 75).

For exploring the relationships between the CADs of the colourants of a given BCT, the richness estimation, and the sample size, let us assume a statistical model in which the chosen BCT is BLUE, the number of already analysed BLUE spots[2] from a MS is 50, and five the number of colourants. Many CADs are possible: Figure 1 shows two ideal extreme distributions: the curve OAB refers to a perfect even assemblage in which each colourant is equally abundant (10 individuals), whereas the uneven assemblage includes four colorants, the abundances of which are equal to one individual, and the abundance of the fifth is 46 individuals.

The *cumulative abundances* (on the horizontal axis) of the two distributions are constrained under the rule of successively sampling at random an individual from each assemblage, without replacement[3]. The richness of both distributions is five, but this data emerges earlier for the even assemblage due to the presence of the asymptotic segment A_B. Instead, the curve of the uneven assemblage is bound to converge by the model's constraints with the even curve at the x-coordinate= 50.

The two curves represent two assemblages with different evenness, which is meant as a community property in which, if each colourant present is equally abundant, then the MS has high evenness, and if the colourants differ widely in abundance, the assemblage has low evenness. It is apparent that actual CADs will be somewhere between the end-types of Fig. 1, and that uneven assemblages require larger sample sizes that even ones. Given a certain level of confidence (e.g. 95 %), the sample size increases exponentially if the rare colourants make up 5% of the sample, thus becoming incompatible with the slot of time available for the analysis of the cultural items. For estimating colourant richness from samples in the case of bio-richness, see Chazdon et al. (1998).

SEVEN PROBLEMS IN THE COURSE OF THE ANALYTICAL PROCESS

#1. *Codicological and palaeographic description of a generic MS*. The codicological and palaeographical studies of a MS provide significant clues on its possible heterogeneity; if homogeneous, the MS represents a statistical unit of analysis consisting of a set of homogeneous illuminations and its correlated text. If heterogeneous, the different parts of the MS are different statistical units.

[2]The colour name of the stained spot may be sapphire-blue, indigo-blue, light blue, dark blue, etc., that is, one of the many colour terms included in the BLUE BCT, see Kerttula (2002: 175-186).

[3]Data is averaged over 50 randomizations, Magurran (2003: 75).

#2. *Choice of a generic stained spot(i) (i=1,2,3…n) of a given folio and MS*, and *ascription of a colour term to the spot(i).* These operations occur at the same time, but they require different cautions. The analysed number of stained spots for a given BCT, or BCT sample size, must comply a certain level of confidence and permit statistical computations, and the ascription of a colour term to each *spot (i)* is a multi step process.

In a preparatory phase, the analyst performs a first judgement of the illuminations using pictures of the illuminations themselves, chooses the stained spots to be analysed, and associates a colour term according to his perception and experience to each of them. Successively, when the analyst has access to the illuminations of the real manuscript, the previously assigned colour terms can be checked, and some errors may be corrected. Later, a BCT is associated to the colour term of the stained spot, and to the results of spectrographic analyses, then it is possible to check the consistency of the entire process, and further mistakes could be fixed. In a scientific context, the visual stimuli which are in most of the cases presented to subjects are calibrated, using Munsell charts. However, unlike archaeological soil studies, where Munsell charts are used, the estimation of a colour term requires less accuracy, as many colour terms are associated with the same BCT.

The distribution of BCTs in the Munsell system can be determined through an ongoing colour naming experiment[4]. Colour stimuli of about 3 degrees in size from a viewing distance of 50cm were presented against a neutral mid-grey background with a black outline of 1 pixel. The 600 test stimuli of the colour naming experiment were selected from the Munsell Renotation Data. Colours were specified in the sRGB standard colour space of the Internet and out of gamut colours were removed. Participation was voluntary and anonymous. For this study, we used a refined dataset from 330 English speaking observers with a normal colour vision producing 5428 responses, see Mylonas and MacDonald (2016). For each BCT responded by the observers of the experiment, we computed the empirical arithmetic mean and variance-covariance matrix in CIELAB assuming D65 illuminant. These Gaussian distributions for each BCT can be used to assign automatically a BCT to spot measurements and or image regions of artworks in CIELAB. Lighting conditions as close as possible to the standard one, which is called D65, theoretically 40 cm from the MS placed horizontally[5]. Average midday light in Western Europe corresponds roughly to D65.

#3. *Ascription of the BCT corresponding to the colour name*: this may be made by using Kerttula's *English colour terms* (2002), which gives a large set of related colour terms for each BCT.

#4. *A codebook for treating data*. The smallest unit of the analytical work is the stained spot(i), namely its analytical features consisting of four sets of pieces of information, which are collected in a vector i.e.:i) The cultural institution, the MS and its production. ii) The co-ordinates and features of the *spot(i)*. iii) Spectrographic data. iv) Invasive analysis: in some cases, it is possible to find detached fragments of painting in the gutter of a book, which cannot be repositioned, and the fragments can be subjected to micro-invasive analysis with powerful techniques such as HPLC-MS or Raman-SERS, in order to have more accurate information on the colourants.

#5. *Increasing number of records involve a SPSS type application*. The main problem of using an applicative of SPSS type is that the list of colourant classes must be clearly defined and closed before entering the records, and must be open to accept analytical results with greater or lesser accuracy. For example, an analyst may not be able to distinguish whether a given red lake is of animal or vegetal origin. In this latter case the applicative should contain a poor but still useful class named 'red lac',

---

[4] Designed by Mylonas and MacDonald (2016), it is available at https://colournaming.org.
[5] https://web.archive.org/web/20171204115107/http://www.cie.co.at/publ/abst/s005.html (retrieved 2021, July 6th).

together with the classes ‘red lac of vegetal origin’, ‘red lac of animal origin’, and the most precise class, for example red lac from *Kerria lacca* (Kerr). Another example, even more common, is the classification of pigments containing iron oxides. There are a lot of pigments in which the chromophore system is due to hematite (Fe_2O_3) or goethite ($FeO{\cdot}OH$). As far as non-invasive techniques are concerned, it is not possible to go beyond the identification of hematite-containing and goethite containing pigment. In the end, the classification of such pigments could limit to more generic terms such as “hematite-containing” or “red iron oxide pigment” and “goethite-containing” or “yellow iron oxide pigment”.

#6. *Diversity between two MS or collections of MSS.* The notion of diversity is used across a big number disparate discipline: the present work aims to compare two sets of colourants used in two collections of MSS. Two different kinds of comparison can be made: the first concerns the presence/absence data, whereas in the second case one considers the relative abundances of the colourants. A legion of diversity indices had been proposed by the literature, see Schroeder and Jenkins (2018). In the present paper we use the simple diversity index DI (usually called segregation index in the literature), which is used in social sciences since the 1950s, see Duncan and Duncan (1955). Let us consider, for example, the two vectors of Table 1, the DI is simply half of the summation of the absolute differences of the pairs of cells of the two vectors, namely $0.5\Sigma|a_i - b_i|=0.52$.

#7. *The features of a photometric procedure for automatic BCTs scanning.* Given a generic miniature, the application aims to evaluate for each BCT the portion of the coloured surface, and therefore the BCTs distribution of the illumination. The procedure requires to:

i) choose a colour space (e.g., CIELAB), and its related metric, which defines a distance operator $d(\mathbf{x_i}, \mathbf{x_j})$ between two colour vectors $\mathbf{x_i}$ and $\mathbf{x_j}$. The choice of colour space $\boldsymbol{S}$ and distance operator $\boldsymbol{d}$ influences the properties of the subsequent clustering algorithm.

ii) Define the eleven centroids $\mathbf{x_0}$ to $\mathbf{x_{11}}$, one for each of the 11 BCTs through the crowdsourcing experimental procedure mentioned in #2.

iii) Define either one distance threshold $\mathbf{K}$ or one distance threshold $\mathbf{K_i}$ for each colour centroid $\mathbf{x_i}$. informed by the covariance matrix of each BCT as determined in step #2. The threshold(s) will be used to assign all vectors within a hypersphere centred around each centroid to that specific colour. A candidate vector $\mathbf{x}$ will be assigned to the colour i if and only if $d(x, \mathbf{x_i}) < K_i$. A colour vector which is not close enough to any of the centroid will be assigned to a null-class.

iv) An algorithm is needed, once the thresholds are set, for scanning the illuminations and classify the colours according to the metric and thresholds.

Further, histograms of the colours distribution for the whole surface can be generated for each MS and aggregated appropriately, and the program should compute the BCT average values of the coloured area of two generic collections of MSS given the distribution of the BCTs and the area examined of each single MS.

N.B.: The parameters $\mathbf{K_i}$ can be learned by applying a supervised learning algorithm to a labelled colour dataset, if this is obtainable with reasonable effort. In alternative, expert knowledge can be used to set the thresholds Note that the procedure described above requires an appropriate colorimetric calibration procedure in order to be reproducible across acquisition devices and samples.

REFERENCES

Berlin, B., and P. Kay. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley and Los Angeles: University of California Press.

Chazdon R. L., R. K. Colwell, J. S. Denslow, and M. R. Guariguata. 1998. Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of North-eastern Costa Rica. In: *Forest biodiversity research, monitoring and modelling: conceptual background and old world case studies*, F. Dallmeier, J.A. Comiskey (eds). Paris: UNESCO, New York: Parthenon, 285-309.

Duncan, O. D., and B. Duncan. 1955. Methodological Analysis of Segregation Indexes. *American Sociological Review* 20, 2: 210-217.

Kerttula, S. 2002. *English colour terms: etymology, chronology, and relative basicness*. Helsinki: Société Néophilologique.

Magurran, E. A. 2004. *Measuring biological diversity*. Malden (MA): Blackwell Publishing

McGill, J. B. 2011. Species abundance distributions. In: *Biological diversity: frontiers in measurement and assessment*, A E. Magurran, and J. B. McGill (eds.). Oxford, New York: Oxford University Press, 105-122.

Mylonas, D., and L. MacDonald. 2016. Augmenting Basic Colour Terms in English. *Color research and application* 41, 1: 32-42.

Schroeder, P. J., and G. D. Jenkins. 2018. How robust are popular beta diversity indices to sampling error? *Ecosphere* 9, 2: e02100. doi:10.1002/ecs2.2100.