An Experimental Comparison of Three Natural Language Colour Naming Models

DAMIAN CONWAY

Victorian Centre for Image Processing and Graphics
Department of Computer Science
Monash University
Clayton, Victoria 3168
Australia

email: damian@cs.monash.edu.au

Abstract: The problems inherent in providing natural language generation of colour names are discussed. Three models for generating natural language descriptions of HSL colours are described. The effectiveness of these models in describing colours is compared experimentally. It is concluded that a rigid syntactic mapping of HSL components to orthogonal linguistic axes is generally inferior to exhaustive enumeration of colours or custom selection of adjectival colour modifiers. Interesting variations of model preference for different hues and different numbers of adjectival modifiers are noted.

Introduction

Research in colour modelling for computational purposes has concentrated on finding numerical representations of colours which are convenient for use in computer graphics and image processing.

These representation schemes use a three dimensional colour space (RGB, CMY, YIQ, HSV, HSL, XYZ, L*a*b*, L*u*v*, etc.) and quantify colours as points in that space [3,4,6,9]. This provides a means of accurately specifying most colours in terms of a small set of parameters which are suitable for subsequent numeric manipulation.

Schwarz, Cowan and Beatty [7] conducted an extensive comparison of human colour matching performance using five standard numerical colour models. Their study indicated that considerable experience is required to effectively understand and use numerical colour specifications.

The problem is that none of these colour spaces map well onto the ways humans think about colours. Given a random set of RGB or HSL values, even experienced humans can have difficulty determining what colour is being represented.

What colour is represented by the RGB triple [1.0 , 0.5 , 0.25]?

What about the normalised HSL triple [0.06 , 0.75 , 1.0]?

The reverse process is perhaps even more difficult.

What is the RGB triple which represents the colour *"khaki"*?

What is the equivalent HSL triple ?

At Monash University we are developing a multi-media interface which coordinates automatically-generated natural language explanations with synthetic animations of information to be conveyed. It is in this context that the two related problems of generation and comprehension of natural language colour descriptions have arisen. For example, it is easier to understand software which refers to a colour as *"light tan"*, rather than *"RGB [1.0 , 0.5 , 0.25]"*. Similarly, people generally prefer to say *"khaki"* , instead of *"HSL [0.2, 0.4 , 0.95]"*

In 1982 Berk, Brownston and Kaufman [1,2] proposed a semantic classification system called CNS (for "Colour Naming System"), based on the widely used ISCC-NBS system [11]. CNS is a simple grammar which linguistically encodes a quantised HSL space of 627 distinct colours. The study suggested that proficient CNS users could more accurately identify colours than equally experienced HSL or RGB users.

This paper proposes two new semantic colour models and compares them with a variant of the CNS model. The intention is to gather evidence regarding the manner in which humans conceive and describe colours they experience and hence draw conclusions regarding the best way to present colours linguistically.

The basis of the proposed models is the semantic classification of colours originally represented using a normalised HSL colour space [9]. The normalised HSL model represents a colour by three real values in the range 0 to 1.

The first value encodes the colour's dominant **hue** (H), progressing cyclically from *red* at 0.0, through *green* at 0.33, *blue* at 0.67 and back to *red* at 1.0. Intermediate values encode composite colours such as *yellow*, *turquoise* and *purple*.

The second value encodes the purity or tone of the colour (its **saturation** S), ranging from no colour (a *grey* tone) at 0.0, to a vivid tone of the colour (no *grey*) at 1.0.

The final value encodes the brightness of the resulting shade (its **luminance** L), from *black* at 0.0 to full brightness at 1.0.

Proposed Colour Models

1. Comparison model

The Comparison model is a direct semantic quantisation of HSL space. Each colour name corresponds to a particular point in HSL colour space and to its immediate neighbourhood (that is: the unnamed colours closer to that named HSL point than to any other named point.) In this model "closeness" is defined as Euclidean distance in normalised HSL space, though any other suitable metric could be used.

The model postulates that humans use a discrete, empirical, non-uniform colour name space, constructed by comparison with their real-world experience of colour. Colour names in this space are semantic labels for the fuzzy sets of real-world phenomena with which colours are compared.

A very small Comparison colour space	
HSL triple	**Colour name**
[0.11 , 0.27 , 0.96]	***wheat***
[0.40 , 0.67 , 0.55]	***sea green***
[0.58 , 0.22 , 0.56]	***slate***
[0.62 , 0.64 , 0.67]	***cobalt***
[0.99 , 0.83 , 0.89]	***alizarin crimson***

Each person's Comparison colour space (CCS) will be unique, reflecting their vocabulary and individual experience of colours in the real world. Colour names in the CCS will be distributed irregularly through HSL space according to the individual.

Some people may have a very extensive CCS and will know and be able to linguistically distinguish hundreds of colours. Others may have a comparatively small CCS, knowing and using only a few dozen individual colour names. Significant variations in CCS are also to be expected between cultures.

It is also possible that many people will have a larger CCS for recognition than for recall. That is, on seeing a colour, a person asked to select the name of the colour from a list may be able to classify the colour with more discrimination than if they were to attempt the task without prompting.

A Comparison colour space can be implemented as a simple list of distinct colour names and associated HSL values. Classification of an arbitrary HSL-specified colour is then simply a process of determining the closest HSL value in the CCS and assigning the associated name to the colour.

Example: Find the CCS name for the colour represented by *HSL [0.3 , 0.3, 0.3]*	
Colour name	**Distance to [0.3,0.3,0.3]**
wheat	***0.68***
sea green	***0.46***
slate	***0.39***
cobalt	***0.59***
alizarin crimson	***1.05***
Hence classify *HSL [0.3 , 0.3, 0.3]* as *"slate"*	

Obviously, the accuracy with which a colour can be classified will be determined by the number and chromatic distribution of colours in the CCS. For the purposes of this experiment a list of 179 colour names was compiled, with a mean Euclidean separation between nearest neighbours of 0.12 (maximum separation 0.73)

2. Qualification model

The Qualification model is a variant implementation of the CNS model. The model hypothesises that most individuals have only a very small Comparison colour space. Alternatively, it suggests that although many people may possess a substantial colour vocabulary, they use only a small subset of it in everyday life.

This smaller set of base colour names, perhaps only 10 to 20 in total, represents a crude quantisation of hue (H value). Classification is refined, where necessary, by the use of appropriate adjectives to discriminate between various tones (S values) and shades (L values) of these hues.

Qualification colour model (examples)	
HSL triple	**Colour name**
[0.11 , 0.27 , 0.96]	***pale bright yellow***
[0.40 , 0.67 , 0.55]	***aqua***
[0.58 , 0.22 , 0.56]	***pale blue***
[0.62 , 0.64 , 0.67]	***blue***
[0.99 , 0.83 , 0.89]	***bright red***

The model assumes that the list of base colour names is likely to vary only slightly within a population, and is likely to correspond to what that population considers as "common colours". Some support for this hypothesis can be seen in the fact that the Macquarie Thesaurus [10] lists only 7 principal categories of colour.

Experiment: Name 10 colours.
Compare your list with the hues in table 1.

The Qualification colour model can be implemented by quantising the hue value of a given HSL colour into the small set of base colour names and then prepending an adjective signifying the saturation value and another signifying the luminance. Note that either or both of these adjectives may be omitted, usually indicating mid-range values. Appendix A gives a complete grammar for this process.

The order of prepending adjectives has semantic significance. The adjective closer to the colour name appears to bind more strongly and exert a more fundamental modification to the hue. However, the adjective that is read first strongly biases subsequent semantic analysis. Thus *"pale bright blue"* may be interpreted very differently from *"bright pale blue"*.

A third complicating factor is that in general, intensity adjectives seem to exert more influence than saturation adjectives, regardless of position. This is perhaps because the concept of luminance is more directly perceived than that of saturation. The subtleties of adjective order in colour semantics are noted here, but not further investigated in this paper.

For the purposes of this experiment, 14 base colours where chosen. Two intensity modifiers were selected and three saturations modifiers. Where saturation or luminance was less than 0.1 or luminance greater that 0.9, all hues where classified as some shade of grey, according to luminance. The resultant colour space consists of 141 distinct colour names. Note that this implementation of CNS differs from Berk et al. in that it uses one fewer saturation and luminance modifiers. Furthermore, so as to differentiate the model from the Sensory model below, these adjectives were chosen so as to be free of obvious emotional connotation.

Table 1 summarises the model. An asterisk indicates a special case (*grey*). A dash indicates a value range for which no modifying adjective is required.

Table 1: Qualification colour model

Value (of S, L, or H)	Saturation adjective	Luminance adjective	Hue base name
< 0.1	*	*	*orange/brown*
0.1 - 0.2	*pale*	*dark*	*yellow*
0.2 - 0.3	*pale*	*dark*	*green*
0.3 - 0.4	*pale*	-	*aqua*
0.4 - 0.5	*pastel*	-	*aquamarine*
0.5 - 0.6	*pastel*	-	*blue*
0.6 - 0.7	-	-	*violet*
0.7 - 0.8	-	-	*purple*
0.8 - 0.9	-	*bright*	*magenta*
> 0.9	*pure*	*	*red*

3. Sensory model

The Sensory model is in some respects a hybrid of the Comparison and Qualification models. The base colour names are the same as those of the Qualification model, but the qualifying adjectives are drawn from a set similar in nature to the colours in the Comparison colour space.

The Sensory model hypothesises that people rarely combine more than one adjective when qualifying a colour and that the adjectives they choose tend to be drawn more frequently from sensory or emotional experience than from a strict table of luminance/saturation modifiers. As a result, colour modifiers tend to be more lyrical and analogous than functional.

Sensory colour model (examples)	
HSL triple	**Colour name**
[0.11 , 0.27 , 0.96]	***pallid yellow***
[0.40 , 0.67 , 0.55]	***dusky green***
[0.58 , 0.22 , 0.56]	***washed-out blue***
[0.62 , 0.64 , 0.67]	***dusky blue***
[0.99 , 0.83 , 0.89]	***soft red***

Hence, though it may be strictly accurate, people tend not to describe a colour as "dark pale blue" and may even consider this a contradiction. This suggests that the adjectives with which people qualify colour descriptions may be drawn from a single, non-uniform, multidimensional continuum, rather than two orthogonal one-dimensional adjectival spaces.

In particular the Sensory model quantises the luminance/saturation plane into a collection of discrete regions, each with a characteristic adjective. These characteristic adjectives are typically epithets transferred from the senses of touch or smell, or from the realm of emotion.

Table 2: Sensory colour model adjectives				
Saturation	**Luminance**			
	< 0.1	***0.1 - 0.3***	***0.3 - 0.8***	***> 0.8***
< 0.1	*	*	*	*
0.1 - 0.4	*	***gloomy***	***washed-out***	***pallid***
0.4 - 0.6	*	***murky***	***dull***	***hazy***
0.6 - 0.9	*	***earthy***	***dusky***	***soft***
> 0.9	*	***sombre***	***subdued***	***clear***

As with the Qualification model, special cases exist when luminance is very low (*black*) or saturation is low (shades of *grey*). In these cases (indicated in Table 2 by an asterisk) the adjective applied depends on the hue. *Greys* which are slightly *red* or *yellow* in hue are labelled *"warm"*; *greys* which tend towards the *blue/green* are labelled *"cool"*.

Experimental procedure

The goal of the research was to determine which, if any, of the three proposed colour naming models best reflected the way individuals name colours. To simplify analysis, the response space was restricted by prompting for selection between fixed alternatives, rather than prompting for a general response. Hence this experiment tested only recognition, not recall of colour names.

The subject population chosen was a random sample of 248 people attending the 1991 Monash University Open Day. Participation was informed and voluntary, with testing conducted by unsupervised interaction with carefully designed software. Precautions were taken however to ensure that very young participants were appropriately supervised.

Subjects were asked to view a sequence of ten randomly generated colours, projected on a black background. The colours where displayed on a 24-bit RGB screen as a rectangle approximately 35cm by 20cm.

As each colour was displayed, subjects were offered a list of three colour names, generated from the HSL values of the colour using each of the three models. The order in which the colour names were listed was randomised to avoid systematic bias. Subjects were asked to select the name which "best" described the colour they were viewing.

Subjects preferences were then classified into one of ten categories. Where 70% or more of a subject's responses corresponded to one particular model, the subject was classified as responding strongly to that model *(StrongC, StrongQ* or *StrongS).* Where 80% or more of the subject's responses were evenly divided between two models, the subject was classified as responding strongly to those two models *(StrongCQ, StrongCS* or *StrongQS).* Where 50% or more of a subject's responses corresponded to one model and no more than 30% of their other responses corresponded to any other model, the subject was classified as showing a slight response to a particular model *(WeakC, WeakQ* or *WeakS).* The remaining subjects where classified as showing no response to any particular model *(Neutral).*

The null hypothesis H_0 was that responses would be drawn from a uniform random distribution. That is, within each classification that no preference would be seen for any of the proposed models. Under H_0 the expected total percentages of population in each classification are:

Expected classification given null hypothesis	
Classification	**Expected percentage**
StrongC, Q or S	***6.0%***
StrongCQ, CS or QS	***32.2%***
WeakC, Q or S	***40.5%***
Neutral	***21.3%***

Observations

Table 3 shows the actual classifications resulting from 248 trials (2480 responses.)

Table 3: Actual classification of subjects	
Classification	**Percentage**
Strong C	***3.2%***
Strong Q	***1.6%***
Strong S	***4.0%***
	8.8%
Strong CQ	***4.8%***
Strong CS	***17.7%***
Strong QS	***6.0%***
	28.5%
Weak C	***12.5%***
Weak Q	***6.0%***
Weak S	***15.7%***
	34.2%
Neutral	***28.5%***

The trend evident in the 71.5% of trials in which subjects indicated some preference for particular colour models is made clearer in Chart 1. Here *Weak* and *Strong* preferences are consolidated. The broken line represents the expected results given the null hypothesis and the surrounding dark shaded region is the uniform 0.01 χ^2 acceptance region for the null hypothesis (that is, the range of values for which the probability that H_0 is valid exceeds 1 percent).

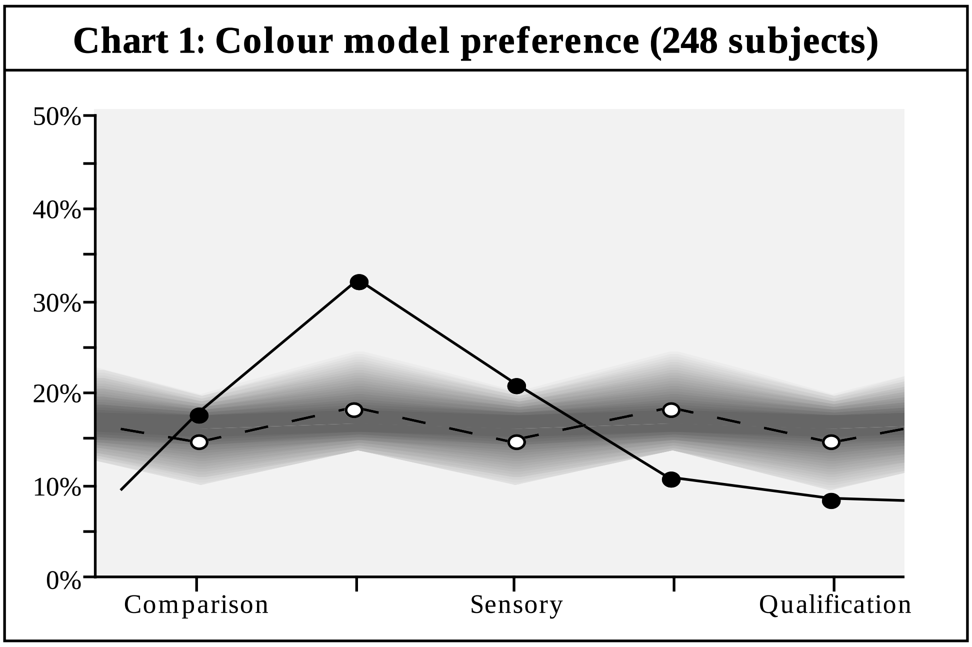
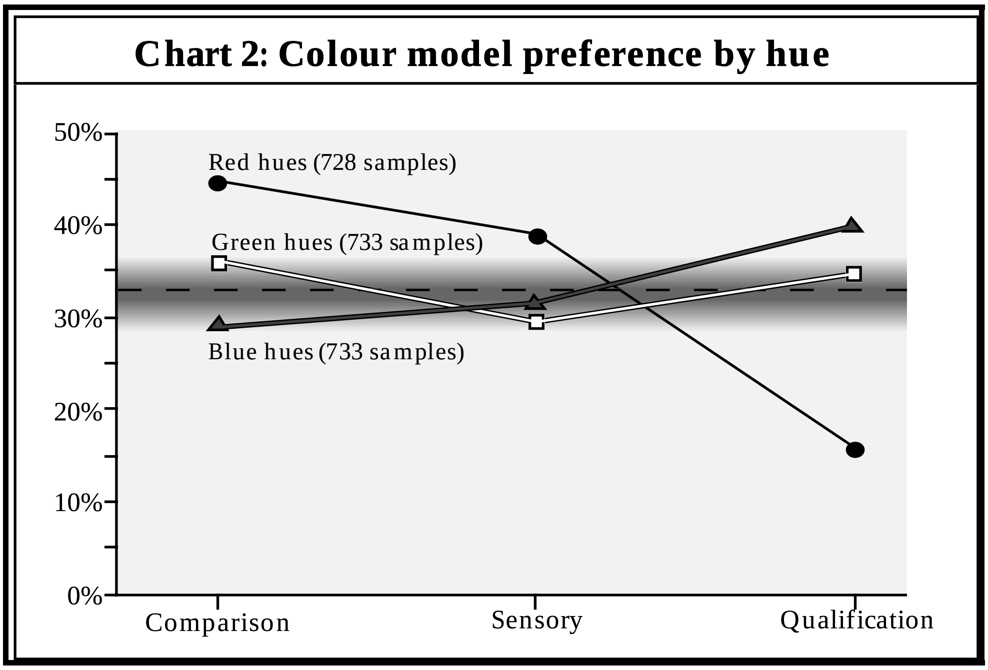


Chart 1 reveals a clear and balanced preference for colour names generated by either the Comparison or Sensory models. The probability that such a preference exists in the data exceeds 0.995 (χ^2 measure).

A strong differentiation between colour model preferences also appears when the data is analysed according to the dominant hue of the colours presented. Chart 2 shows the relative frequency with which colour names generated by a particular model were selected, as a function of the dominant hue of the colour displayed. Here classification is on a per sample basis, rather than a per subject basis. Once again the broken line indicates the null hypothesis and the dark shaded region the 0.01 acceptance range for H_0.



For *reddish* hues, the data reflects the previously noted strong preference for the Comparison and Sensory models. Names for *green* hues are selected more or less uniformly from any of the three models. However, when presented with *bluish* colours, the subjects showed a slight preference for the Qualification model.

As hypothesised in the Sensory model, there is a clear disinclination amongst subjects to use more than one adjective to qualify a colour. Analysis of subject reactions to the Qualification model shows a strong correlation between the number of adjectives used to qualify a colour name and the probability that the name proposed by the Qualification model would be selected.

Colour names which contained either a single luminance modifier (for example: *"dark..."*) or two modifiers (for example: *"light pastel..."*) had a lower probability of selection than names with no modifiers or a single saturation modifier (for example: *"pale..."*). Table 4 summarises these results.

Table 4: Adjective count in Qualification model

Adjectives	Number of cases	Probability of selection
None	***196***	***0.28***
Saturation only	***121***	***0.31***
Luminance only	***141***	***0.22***
Both	***98***	***0.22***

Implications

It is clear that a marked preference for colour descriptions generated using the Comparison and Sensory models existed in the subject population. However, when classifying colours of predominantly primary hue, this preference was only statistically significant when classifying *reddish* hues. Names for *green* hues were found to be approximately equally satisfactory whatever model was used to generate them. The subjects also displayed a slight preference for names generated using the Qualification model when classifying *bluish* hues.

It is speculated that the neutral response to green hues may reflect the prevalence of that part of the spectrum in a wide variety of everyday human experiences. The response to *bluish* colours may stem from the known insensitivity of the human eye to that portion of the spectrum [8] leading to a preference for the more "obvious" colour names generated by the Qualification model.

Although the Qualification model was least favoured, this bias effectively disappeared when adjectival modification was omitted or restricted to a single saturation modifier. In such cases, response to the Qualification model was consistent with the null hypothesis.

This analysis would suggest that for the purpose of presenting colours linguistically, a hybrid Sensory-Comparison model is to be preferred. The

Comparison colour space should be moderately dense (say several hundred names) and consist of Sensory-style modifications of some small set of hues. No more than one adjective should be applied to each colour name. If a luminance and saturation modifier appear equally applicable to a particular colour, the saturation modifier should be chosen.

References

[1] Berk, T., L. Brownston, and A. Kaufman, *A new color-naming system for graphics languages*, IEEE CG&A, vol. 2, no. 3, pp 37-44, IEEE, May 1982

[2] Berk, T., L. Brownston, and A. Kaufman, *A human factors study of color notation systems for computer graphics,* Comm. ACM, vol. 25, no. 8, pp 547-550, ACM, August 1982

[3] Commission Internationale de l'Eclairage, *CIE Proceedings 1931,* Cambridge University, 1932.

[4] Commission Internationale de l'Eclairage, *Recommendations on Uniform Colour Spaces, Color-Difference Equations, Psychrometric Colour Terms."*Bureau Central de la CIE, 1978

[5] Feynman, R. P., R. B. Leighton, and M. Sands, *Color Vision*, The Feynman Lectures on Physics, vol. 1, Addison-Wesley, 1963.

[6] Hall, R., *Illumination and Color in Computer Generated Imagery,* Springer-Verlag, NY, 1989

[7] Joblove, G. H., and D. Greenberg, *Color Spaces for Computer Graphics,* Computer Graphics, vol. 12, no. 3, pp. 20-25, ACM, August 1978.

[8] Schwarz, M. W., W. B. Cowan, and J. C. Beatty, *An Experimental Comparison of RGB, YIQ, LAB, HSV, and Opponent Color Models,* ACM TOG, vol. 6, no. 2, pp 123-158, ACM, April 1987

[9] Smith, A. R., *Color Gamut Transform Pairs,* Computer Graphics, vol. 12, no. 3, pp. 12-19, ACM, August 1978.

[10] *The Macquarie Thesaurus,* edited by J.R.L. Bernard, pp. 126.3-128.2, Penguin Books, Melbourne, 1986

[11] US Department of Commerce, National Bureau of Standards, *Color: Universal Language and Dictionary of Names,* NBS special publication 440, US Government Printing Office, Washington D.C., S.D. Catalog No. C13.10:440, 1976

Appendix A: Generation of colour names

This appendix provides grammars for the Qualification and Sensory models[†] which may used to generate or recognize colour names. Terminals appear in a fixed-width font thus: **`terminal`**. Comments, which appear in italics, indicate the range for which particular substitutions apply (the symbols H, S and L refer to the corresponding components of the normalized HSL representation of the colour.) Note that the rewriting rules for Tone and Hue are common to both models and hence are presented only in the first grammar.

1. Qualification model

Colour	←	Luminance Tone	*iff* $0.0 \le L < 0.1$ *or* $0.0 \le S < 0.1$
	←	Luminance Saturation Hue	*otherwise*
Tone	←	**`black`**	*iff* $0.0 \le L < 0.1$
	←	**`grey`**	*iff* $0.1 \le L < 0.9$
	←	**`white`**	*iff* $0.9 \le L \le 1.0$
Luminance	←	**`dark`**	*iff* $0.1 \le L < 0.3$
	←	**`bright`**	*iff* $0.8 \le L < 0.9$
	←	Nothing	*otherwise*
Saturation	←	**`pale`**	*iff* $0.1 \le S < 0.4$
	←	**`pastel`**	*iff* $0.4 \le S < 0.6$
	←	**`pure`**	*iff* $0.9 < S \le 1.0$
	←	Nothing	*otherwise*
Hue	←	**`brown`**	*iff* $0.0 \le H < 0.1$ *and* $0.0 \le L < 0.5$
	←	**`orange`**	*iff* $0.0 \le H < 0.1$ *and* $0.5 \le L \le 1.0$
	←	**`yellow`**	*iff* $0.1 \le H < 0.2$
	←	**`green`**	*iff* $0.2 \le H < 0.3$
	←	**`aqua`**	*iff* $0.3 \le H < 0.4$
	←	**`aquamarine`**	*iff* $0.4 \le H < 0.5$
	←	**`blue`**	*iff* $0.5 \le H < 0.6$
	←	**`violet`**	*iff* $0.6 \le H < 0.7$
	←	**`purple`**	*iff* $0.7 \le H < 0.8$
	←	**`magenta`**	*iff* $0.8 \le H < 0.9$
	←	**`red`**	*iff* $0.9 \le H \le 1.0$

2. Sensory model

Colour	←	Tone	*iff* $0.0 \le S < 0.1$ *and* $0.0 \le L < 0.1$
	←	Modifer Tone	*iff* $0.0 \le S < 0.1$ *and* $0.1 \le L \le 1.0$
	←	Adjective Hue	*otherwise*
Modifer	←	**`cool`**	*iff* $0.2 \le H < 0.7$
	←	**`warm`**	*otherwise*
Adjective	←	**`gloomy`**	*iff* $0.1 \le L < 0.3$ *and* $0.1 \le S < 0.4$
	←	**`murky`**	*iff* $0.1 \le L < 0.3$ *and* $0.4 \le S < 0.6$
	←	**`earthy`**	*iff* $0.1 \le L < 0.3$ *and* $0.6 \le S < 0.9$
	←	**`sombre`**	*iff* $0.1 \le L < 0.3$ *and* $0.9 \le S \le 1.0$
	←	**`washed-out`**	*iff* $0.3 \le L < 0.8$ *and* $0.1 \le S < 0.4$
	←	**`dull`**	*iff* $0.3 \le L < 0.8$ *and* $0.4 \le S < 0.6$
	←	**`dusky`**	*iff* $0.3 \le L < 0.8$ *and* $0.6 \le S < 0.9$
	←	**`subdued`**	*iff* $0.3 \le L < 0.8$ *and* $0.9 \le S \le 1.0$
	←	**`pallid`**	*iff* $0.8 \le L \le 1.0$ *and* $0.1 \le S < 0.4$
	←	**`hazy`**	*iff* $0.8 \le L \le 1.0$ *and* $0.4 \le S < 0.6$
	←	**`soft`**	*iff* $0.8 \le L \le 1.0$ *and* $0.6 \le S < 0.9$
	←	**`clear`**	*iff* $0.8 \le L \le 1.0$ *and* $0.9 \le S \le 1.0$

[†] There is no grammar for the Comparison model. It is a look-up scheme.