Colourful language : Measuring word–colour associations
Mohammad, Saif M.

Publisher’s version / Version de l'éditeur:

Proceedings of the ACL 2011 Workshop on Cognitive Modeling and Computational Linguistics, 2011-07-01

Colourful Language: Measuring Word–Colour Associations

Saif Mohammad
Institute for Information Technology
National Research Council Canada
Ottawa, Ontario, Canada, K1A 0R6
saif.mohammad@nrc-cnrc.gc.ca

Abstract

Since many real-world concepts are associated with colour, for example *danger* with red, linguistic information is often complimented with the use of appropriate colours in information visualization and product marketing. Yet, there is no comprehensive resource that captures concept–colour associations. We present a method to create a large word–colour association lexicon by crowdsourcing. We focus especially on abstract concepts and emotions to show that even though they cannot be physically visualized, they too tend to have strong colour associations. Finally, we show how word–colour associations manifest themselves in language, and quantify usefulness of co-occurrence and polarity cues in automatically detecting colour associations.[1]

1 Introduction

Colour is a vital component in the successful delivery of information, whether it is in marketing a commercial product (Sable and Akcay, 2010), designing webpages (Meier, 1988; Pribadi et al., 1990), or visualizing information (Christ, 1975; Card et al., 1999). Since real-world concepts have associations with certain colour categories (for example, *danger* with red, and *softness* with pink), complimenting linguistic and non-linguistic information with appropriate colours has a number of benefits, including: (1) strengthening the message (improving semantic coherence), (2) easing cognitive load on the receiver, (3) conveying the message quickly, and (4) evoking the desired emotional response. Consider, for example, the use of red in stop signs. Drivers are able to recognize the sign faster, and it evokes a subliminal emotion pertaining to danger, which is entirely appropriate in the context. The use of red to show areas of high crime rate in a visualization is another example of good use of colour to draw emotional response. On the other hand, improper use of colour can be more detrimental to understanding than using no colour (Marcus, 1982; Meier, 1988).

Most languages have expressions involving colour, and many of these express sentiment. Examples in English include: *green with envy*, *blue blood* (an aristocrat), *greener pastures* (better avenues), *yellow-bellied* (cowardly), *red carpet* (special treatment), and *looking through rose-tinted glasses* (being optimistic). Further, new expressions are continually coined, for example, *grey with uncertainty* from Bianca Marsden's poem *Confusion*.[2] Thus, knowledge of concept–colour associations may also be useful for automatic natural language systems such as textual entailment, paraphrasing, machine translation, and sentiment analysis.

A word has strong association with a colour when the colour is a salient feature of the concept the word refers to, or because the word is related to a such a concept. Many concept–colour associations, such as *swan* with white and *vegetables* with green, involve physical entities. However, even abstract notions and emotions may have colour as-

[1] This paper is an extended, non-archival, version of the short paper—Mohammad (2011). It provides additional details on the analysis of crowdsourced data, and experiments on the manifestations of word–colour associations in WordNet and in text. It also proposes a polarity-based automatic method.

[2] http://www.biancaday.com/confusion.html

sociations (*honesty*–white, *danger*–red, *joy*–yellow, *anger*–red). Further, many associations are culture-specific (Gage, 1969; Chen, 2005). For example, *prosperity* is associated with red in much of Asia.

Unfortunately, there exists no lexicon with any significant coverage that captures concept–colour associations, and a number of questions remain unanswered, such as, the extent to which humans agree on these associations, and whether physical concepts are more likely to have a colour association than abstract ones. We expect that the word–colour associations manifest themselves as co-occurrences in text and speech, but there have been no studies to show the extent to which words co-occur more with associated colours than with other colours.

In this paper, we describe how we created a large word–colour association lexicon by crowdsourcing with effective quality control measures (Section 3). We used a word-choice question to guide the annotators toward the desired senses of the target words, and also to determine if the annotators know the meanings of the words.

We conducted several experiments to measure the consensus in word–colour associations, and how these associations manifest themselves in language. Specifically, we show that:

- More than 30% of terms have a strong colour association (Sections 4).
- About 33% of thesaurus categories have strong colour associations (Section 5).
- Abstract terms have colour associations almost as often as physical entities do (Section 6).
- There is a strong association of emotions and polarities with colours (Section 7).
- Word-colour association manifests itself as closeness in WordNet (to a smaller extent), and as high co-occurrence in text (to a greater extent) (Section 8).

Finally, we present an automatic method to determine word–colour association that relies on co-occurrence and polarity cues, but no labeled information of word–colour associations. It obtains an accuracy of more than 60%. Comparatively, the random choice and most-frequent class supervised baselines obtain only 9.1% and 33.3%, respectively. Such approaches can be used to for creating similar lexicons in other languages.

2 Related Work

The relation between language and cognition has received considerable attention over the years, mainly on answering whether language impacts thought, and if so, to what extent. Experiments with colour categories have been used both to show that language has an effect on thought (Brown and Lenneberg, 1954; Ratner, 1989) and that it does not (Bornstein, 1985). However, that line of work does not explicitly deal with word–colour associations. In fact, we did not find any other academic work that gathered large word–colour associations. There is, however, a commercial endeavor—Cymbolism[3].

Child et al. (1968), Ou et al. (2011), and others show that people of different ages and genders have different colour preferences. (See also the online study by Joe Hallock[4].) In this work, we are interested in identifying words that have a strong association with a colour due to their meaning; associations that are not affected by age and gender preferences.

There is substantial work on inferring the emotions evoked by colour (Luscher, 1969; Xin et al., 2004; Kaya, 2004). Strapparava and Ozbal (2010) compute corpus-based semantic similarity between emotions and colours. We combine the word–colour and word–emotion association lexicons to determine the correlation between emotion-associated words and colours.

Berlin and Kay (1969), and later Kay and Maffi (1999), showed that often colour terms appeared in languages in certain groups. If a language has only two colour terms, then they are white and black. If a language has three colour terms, then they are white, black, and red. If a language has four colour terms, then they are white, black, red, and green, and so on up to eleven colours. From these groupings, the colours can be ranked as follows:

1. white, 2. black, 3. red, 4. green, 5. yellow, 6. blue, 7. brown, 8. pink, 9. purple, 10. orange, 11. grey (1)

We will refer to the above ranking as the *Berlin and Kay (B&K) order*. There are hundreds of different words for colours.[5] To make our task feasible,

[3]http://www.cymbolism.com/about

[4]http://www.joehallock.com/edu/COM498/preferences.html

[5]See http://en.wikipedia.org/wiki/List_of_colors

we needed to choose a relatively small list of basic colours. We chose to use the eleven basic colour words of Berlin and Kay (1969).

The MRC Psycholinguistic Database (Coltheart, 1981) has, among other information, the *imageability ratings* for 9240 words.[6] The imageability rating is a score given by human judges that reflects how easy it is to visualize the concept. It is a scale from 100 (very hard to visualize) to 700 (very easy to visualize). We use the ratings in our experiments to determine whether there is a correlation between imageability and strength of colour association.

3 Crowdsourcing

Amazon's Mechanical Turk (AMT) is an online crowdsourcing platform that is especially well suited for tasks that can be done over the Internet through a computer or a mobile device.[7] It is already being used to obtain human annotation on various linguistic tasks (Snow et al., 2008; Callison-Burch, 2009). However, one must define the task carefully to obtain annotations of high quality. Several checks must be placed to ensure that random and erroneous annotations are discouraged, rejected, and re-annotated.

We used Mechanical Turk to obtain word–colour association annotations on a large-scale. Each task is broken into small independently solvable units called *HITs (Human Intelligence Tasks)* and uploaded on the Mechanical Turk website. The people who provide responses to these HITs are called *Turkers*. The annotation provided by a Turker for a HIT is called an *assignment*.

We used the *Macquarie Thesaurus* (Bernard, 1986) as the source for terms to be annotated. Thesauri, such as the *Roget's* and *Macquarie*, group related words into categories. The *Macquarie* has about a thousand categories, each having about a hundred or so related terms. Each category has a *head word* that best represents the words in it. The categories can be thought of as coarse senses or concepts (Yarowsky, 1992). If a word is ambiguous, then it is listed in more than one category. Since a word may have different colour associations when used in different senses, we obtained annotations at word-sense level. We chose to annotate words that had one to five senses in the *Macquarie Thesaurus* and occurred frequently in the *Google N-gram Corpus*. We annotated more than 10,000 of these word–sense pairs by creating HITs as described below.

Each HIT has a set of questions, all of which are to be answered by the same person. We requested annotations from five different Turkers for each HIT. (A Turker cannot attempt multiple assignments for the same term.) A complete HIT is shown below:

Q1. Which word is closest in meaning to *sleep*?

- *car*
- *tree*
- *nap*
- *olive*

Q2. What colour is associated with *sleep*?

- black
- blue
- brown
- green
- grey
- orange
- purple
- pink
- red
- white
- yellow

Q1 is a word-choice question generated automatically by taking a near-synonym from the thesaurus and random distractors. The near-synonym also guides the annotator to the desired sense of the word. Further, it encourages the annotator to think clearly about the target word's meaning; we believe this improves the quality of the annotations in Q2. If a word has multiple senses, that is, it is listed in more than one thesaurus category, then separate questionnaires are generated for each sense. Thus we obtain colour associations at a word-sense level.

If an annotator answers Q1 incorrectly, then we discard information obtained from both Q1 and Q2. Thus, even though we do not have correct answers to Q2, likely incorrect annotations are filtered out. About 10% of the annotations were discarded because of an incorrect answer to Q1. Terms with less than three valid annotations were removed from further analysis. Each of the remaining terms had, on average, 4.45 distinct annotations.

The colour options in Q2 were presented in random order. Observe that we do not provide a "not associated with any colour" option. This encourages colour selection even if the annotator felt the association was weak. If there is no association between a word and a colour, then we expect low agreement amongst the annotators. The survey was approved by the ethics board at the authors' institution.

[6] http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm

[7] Mechanical Turk: www.mturk.com

	white	black	red	green	yellow	blue	brown	pink	purple	orange	grey
overall	11.9	12.2	11.7	12.0	11.0	9.4	9.6	8.6	4.2	4.2	4.6
voted	22.7	18.4	13.4	12.1	10.0	6.4	6.3	5.3	2.1	1.5	1.3

Table 1: Percentage of terms marked as being associated with each colour.

4 Word–Colour Association

The information from multiple annotators was combined by taking the majority vote, resulting in a lexicon of 8,813 entries. Each entry contains a unique word–synonym pair (from Q1), majority-voted colour, and a confidence score—number of votes for the colour / number of total votes. (For the analyses in the rest of the paper, ties were broken by picking one colour at random.) A separate version of the lexicon that includes entries for all of the valid annotations by each of the annotators is also available.[8]

The first row, *overall*, in Table 1 shows the percentage of times different colours were associated with the target term. The second row, *voted*, shows percentages after taking a majority vote from multiple annotators. Observe that even though the colour options were presented in random order, the order of the most frequently associated colours is identical to the Berlin and Kay order (Section 2:(1)).

Table 2 shows how often the size of the majority class in colour associations is one, two, three, four, and five. Since the annotators were given eleven colour options to choose from, if we assume independence, then the chance that none of the five annotators agrees with each other (majority class size of one) is $1 \times 10/11 \times 9/11 \times 8/11 \times 7/11 = 0.344$. Thus, if there was no correlation among any of the terms and colours, then 34.4% of the time none of the annotators would have agreed. However, this happens only 15.1% of the time. A large number of terms have a majority class size $\geq$ 2 (84.9%), and thus more than chance association with a colour. One can argue that terms with a majority class size $\geq$ 3 (32%) have *strong* colour associations.

Below are some reasons why agreement values are much lower than those obtained for certain other tasks, for example, part of speech tagging:

- The annotators were not given a "not associated with any colour" option. Low agreement for certain instances is an indicator that these words have weak, if any, colour association.
- Words are associated with colours to different degrees. Some words may be associated with more than one colour in comparable degrees, and there might be higher disagreement for such instances.
- The target word is presented out of context. We expect higher agreement if we provided words in particular contexts, but words can occur in innumerable contexts, and annotating too many instances of the same word is costly.

[8]Please contact the author to obtain a copy of the lexicon.

majority class size						
one	two	three	four	five	$\geq$ two	$\geq$ three
15.1	52.9	22.4	7.3	2.1	84.9	32.0

Table 2: Percentage of terms in different majority classes.

Nonetheless, the term–colour association lexicon is useful for downstream applications because any of the following strategies may be employed: (1) choosing colour associations from only those instances with high agreement, (2) assuming low-agreement terms have no colour association, (3) determining colour association of a category through information from many words, as described in the next section.

5 Category–Colour Association

Words within a thesaurus category may not be strongly associated with any colour, or they may each be associated with many different colours. We now describe experiments to determine whether there exist categories where the semantic coherence carries over to a strong common association with one colour.

We determine the strength of colour association of a category by first determining the colour c most associated with the terms in it, and then calculating the ratio of the number of times a word from the category is associated with c to the number of words in the category associated with any colour. Only cate-

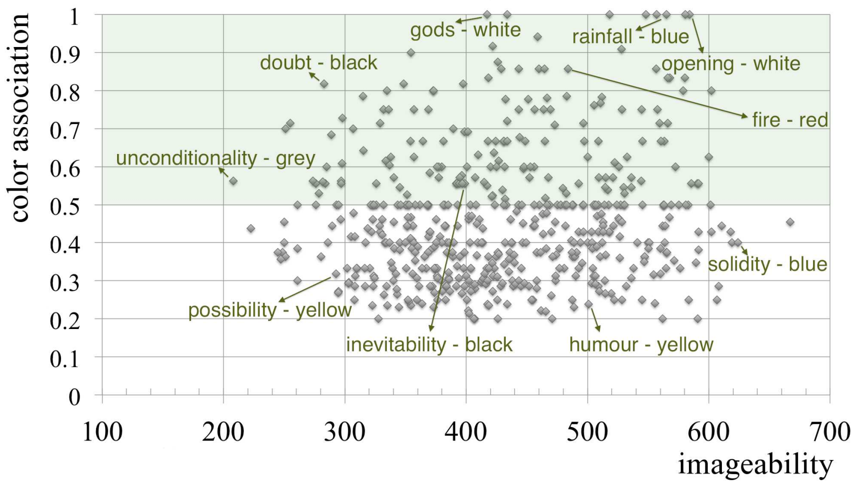

Figure 1: Scatter plot of thesaurus categories. The area of high colour association is shaded. Some points are labeled.

gories that had at least four words that also appear in the word–colour lexicon were considered; 535 of the 812 categories from *Macquarie Thesaurus* met this condition.

If a category has exactly four words that appear in the colour lexicon, and if all four words are associated with different colours, then the category has the lowest possible strength of colour association—0.25 (1/4). 19 categories had a score of 0.25. No category had a score less than 0.25. Any score above 0.25 shows more than random chance association with a colour. There were 516 such categories (96.5%). 177 categories (33.1%) had a score 0.5 or above, that is, half or more of the words in these categories are associated with one colour. We consider these to be strong associations, and a gold standard for automatic measures of association.

6 Imageability and Colour Association

It is natural for physical entities of a certain colour to be associated with that colour. However, abstract concepts such as *danger* and *excitability* are also associated with colours—red and orange, respectively. Figure 1 displays an experiment to determine whether there is a correlation between imageability and association with colour.

We define imageability of a thesaurus category to be the average of the imageability ratings of words in it. We calculated imageability for the 535 categories described in the previous section using only the words that appear in the colour lexicon. Figure 1 shows the scatter plot of these categories on the imageability and strength of colour association axes. The colour association was calculated as described in the previous section.

If higher imageability correlated with greater tendency to have a colour association, then we would see most of the points along the diagonal moving up from left to right. Instead, we observe that the strongly associated categories (points in the shaded region) are spread across the imageability axis, implying that there is only weak, if any, correlation between imageability and strength of association with colour. Imageability and colour association have a Pearson's product moment correlation of 0.116, and a Spearman rank order correlation of 0.102.

7 The Colour of Emotion Words

Emotions such as joy and anger are abstract concepts dealing with one's psychological state. Mohammad and Turney (2010) created a crowdsourced term–emotion association lexicon consisting of associations of over 10,000 word-sense pairs with eight emotions—joy, sadness, anger, fear, trust, disgust, surprise, and anticipation—argued to be the basic and prototypical emotions (Plutchik, 1980). We combine their term–emotion association lexicon and our term–colour lexicon to determine the colour signature of different emotions—the rows in Table 3. The top two most frequently associated colours with each of the eight emotions are shown in bold. For example, the "anger" row shows the percentage of

	white	black	red	green	yellow	blue	brown	pink	purple	orange	grey
anger words	2.1	**30.7**	**32.4**	5.0	5.0	2.4	6.6	0.5	2.3	2.5	9.9
anticipation words	**16.2**	7.5	11.5	**16.2**	10.7	9.5	5.7	5.9	3.1	4.9	8.4
disgust words	2.0	**33.7**	**24.9**	4.8	5.5	1.9	9.7	1.1	1.8	3.5	10.5
fear words	4.5	**31.8**	**25.0**	3.5	6.9	3.0	6.1	1.3	2.3	3.3	11.8
joy words	**21.8**	2.2	7.4	**14.1**	13.4	11.3	3.1	11.1	6.3	5.8	2.8
sadness words	3.0	**36.0**	**18.6**	3.4	5.4	5.8	7.1	0.5	1.4	2.1	16.1
surprise words	11.0	13.4	**21.0**	8.3	**13.5**	5.2	3.4	5.2	4.1	5.6	8.8
trust words	**22.0**	6.3	8.4	14.2	8.3	**14.4**	5.9	5.5	4.9	3.8	5.8

Table 3: Colour signature of emotive terms: percentage of terms associated with each colour. For example, 32.4% of the anger terms are associated with red. The two most associated colours are shown in bold.

	white	black	red	green	yellow	blue	brown	pink	purple	orange	grey
negative	2.9	**28.3**	**21.6**	4.7	6.9	4.1	**9.4**	1.2	2.5	3.8	**14.1**
positive	**20.1**	3.9	8.0	**15.5**	**10.8**	**12.0**	4.8	**7.8**	**5.7**	**5.4**	5.7

Table 4: Colour signature of positive and negative terms: percentage terms associated with each colour. For example, 28.3% of the negative terms are associated with black. The highest values in each column are shown in bold.

anger terms associated with different colours.

We see that all of the emotions have strong associations with certain colours. Observe that anger is associated most with red. Other negative emotions—disgust, fear, sadness—go strongest with black. Among the positive emotions: anticipation is most frequently associated with white and green; joy with white, green, and yellow; and trust with white, blue, and green. Thus, colour can add to the emotional potency of visualizations.

The Mohammad and Turney (2010) lexicon also has associations with positive and negative polarity. We combine these term–polarity associations with term–colour associations to show the colour signature for positive and negative terms—the rows of Table 4. We observe that some colours tend to, more often than not, have strong positive associations (white, green, yellow, blue, pink, and orange), whereas others have strong negative associations (black, red, brown, and grey).

8 Manifestation of Concept–Colour Association in WordNet and in Text

8.1 Closeness in WordNet

Colour terms are listed in WordNet, and interestingly, they are fairly ambiguous. Therefore, they can be found in many different synsets (see Table 5). A casual examination of WordNet reveals that some synsets (or concepts) are close to their associated colour's synset. For example, *darkness* is a hypernym of black and *inflammation* is one hop away from red. It is plausible that if a concept is strongly associated with a certain colour, then such concept–colour pairs will be close to each other in a semantic network such as WordNet. If so, the semantic closeness of a word with each of the eleven basic colours in WordNet can be used to automatically determine the colour most associated with the 177 thesaurus categories from the gold standard described in Section 5 earlier. We determine closeness using two similarity measures—Jiang and Conrath (1997) and Lin (1997)—and two relatedness measures—Lesk (Banerjee and Pedersen, 2003) and gloss vector overlap (Pedersen et al., 2004)—from the WordNet Similarity package.

For each thesaurus category–colour pair, we summed the WordNet closeness of each of the terms in the category to the colour. The colour with the highest sum is chosen as the one closest to the thesaurus category. Section (c) and section (d) of Table 8.2, show how often the closest colours are also the colours most associated with the gold standard categories. Section (a) lists some unsupervised baselines. Random-choice baseline is the score obtained when a colour is chosen at random (1/11 = 9.1%). Another baseline is a system that always chooses the most frequent colour in a corpus. Section (a) reports three such baseline scores obtained by choosing the most frequently occurring colour in three separate corpora. Section (b) lists a supervised baseline obtained by choosing the colour most commonly asso-

colour	white	black	red	green	yellow	blue	brown	pink	purple	orange	grey
# of senses	25	22	7	14	8	16	8	7	7	6	13

Table 5: The number of senses of colour terms in WordNet.

		white	black	red	green	yellow	blue	brown	pink	purple	orange	grey	ρ
	B&K rank:	1	2	3	4	5	6	7	8	9	10	11	
BNC	freq:	1480	3460	2070	1990	270	1430	1170	450	180	360	800	
	rank:	4	1	2	3	10	5	6	8	11	9	7	0.727
GNC	freq:	205	239	138	106	80	123	63	41	16	36	18	
	rank:	2	1	3	5	6	4	7	8	11	9	10	0.884
GBC	freq:	233	188	130	86	44	75	72	14	11	19	22	
	rank:	1	2	3	4	7	5	6	9	10	11	8	0.918

Table 6: Frequency and ranking of colour terms per 1,000,000 words in the *British National Corpus (BNC), Google N-gram Corpus (GNC),* and *Google Books Corpus (GBC)*. The last column lists the Spearman rank order correlation (ρ) of the rankings with the Berlin and Kay (B&K) ranks.

ciated with a categories in the gold standard. The automatic measures listed in sections (c) through (f) do not have access to this information.

Observe that the relatedness measures are markedly better than the similarity measures at identifying the true associated colour. Yet, for a majority of the thesaurus categories the closest colour in WordNet is not the most associated colour.

8.2 Co-occurrence in Text

Physical entities that tend to have a certain colour tend to be associated with that colour. For example leaves are associated with green. Intuition suggests that these entities will co-occur with the associated colours more often than with any other colour. As language has expressions such as *green with envy* and *feeling blue*, we also expect that certain abstract notions, such as *envy* and *sadness*, will co-occur more often with their associated colours, green and blue respectively, more often than with any other colour. We now describe experiments to determine the extent to which target concepts co-occur in text most often with their associated colours.

We selected three corpora to investigate occurrences of colour terms: the *British National Corpus (BNC)* (Burnard, 2000), the *Google N-gram Corpus (GNC)*, and the *Google Books Corpus (GBC)* (Michel et al., 2011).[9] The *BNC*, a 100 million word corpus, is considered to be fairly balanced with text from various domains. The *GNC* is a trillion-word web coprus. The *GBC* is a digitized version of about 5.2 million books, and the English portion has about 361 billion words. The *GNC* and *GBC* are distributed as collections of 1-gram to 5-gram files.

Table 6 shows the frequencies and ranks of the eleven basic colour terms in the *BNC* and the unigram files of *GNC* and *GBC*. The ranking is from the most frequent to the least frequent colour in the corpus. The last column lists the Spearman rank order correlation (ρ) of the rankings with the Berlin and Kay ranks (1969) (listed in Section 2:(1)). Observe that order of the colours from most frequent to least frequent in the *GNC* and *GBC* have a strong correlation with the order proposed by Berlin and Kay, especially so for the rankings obtained from counts in the *Google Books Corpus*.

For each of the 177 gold standard thesaurus categories, we determined the conditional probability of co-occurring with different colour terms in the *BNC, GNC,* and *GBC*. The total co-occurrence frequency of a category with a colour was calculated by summing up the co-occurrence frequency of each of the terms in it with the colour term. We used a four-word window as context. The counts from *GNC* and *GBC* were determined using the fivegram files. Section (e) in Table 8.2 shows how often the colour with the highest conditional probability is also the colour most associated with a category. These numbers are higher than the baselines (a and b), as well as the scores obtained by the WordNet approaches (c).

From Table 5 in Section 7, we know that some

[9] The *BNC* is available at: http://www.natcorp.ox.ac.uk.
The *GNC* is available through the Linguistic Data Consortium.
The *GBC* is available at http://ngrams.googlelabs.com/datasets.

Automatic method for choosing colour	Accuracy
(a) Unsupervised baselines:	
- randomly choosing a colour	9.1
- most frequent colour in *BNC* (black)	23.2
- most frequent colour in *GNC* (black)	23.2
- most frequent colour in *GBC* (white)	33.3
(b) Supervised baseline:	
- colour most often associated with categories (white)	33.3
(c) WordNet similarity measures:	
- Jiang Conrath measure	15.7
- Lin's measure	15.7
(d) WordNet relatedness measures:	
- Lesk measure	24.7
- gloss vector measure	28.6
(e) Co-occurrence in text:	
- $p(colour \mid word)$ in *BNC*	31.4
- $p(colour \mid word)$ in *GNC*	37.9
- $p(colour \mid word)$ in *GBC*	38.3
(f) Co-occurrence and polarity:	
- $p(colour \mid word, polarity)$ in *BNC*	51.4
- $p(colour \mid word, polarity)$ in *GNC*	47.6
- $p(colour \mid word, polarity)$ in *GBC*	**60.1**

Table 7: Percentage of times the colour chosen by automatic method is also the colour identified by annotators as most associated to a thesaurus category.

colours tend to be strongly positive and others negative. We wanted to determine how useful these polarity cues can be in identifying the colour most associated with a category. We used the automatically generated Macquarie Semantic Orientation Lexicon (MSOL) (Mohammad et al., 2009) to determine if a thesaurus category is positive or negative.[10] A category is marked as negative if it has more negative words than positive, otherwise it is marked as positive. If a category is positive, then co-occurrence cues were used to select a colour from only the positive colours (white, green, yellow, blue, pink, and orange), whereas if a category is negative, then co-occurrence cues select from only the negative colours (black, red, brown, and grey). Section (f) of Table 8.2 provides results with this method. Observe that these numbers are a marked improvement over Section (e) numbers, suggesting that polarity cues can be very useful in determining concept–colour association.

[10]MSOL is available at http://www.umiacs.umd.edu/∼saif/WebPages/ResearchInterests.html#semanticorientation.

Counts from the *GNC* yielded poorer results compared to the much smaller *BNC*, and the somewhat smaller *GBC* possibly because frequency counts from *GNC* are available only for those n-grams that occur at least thirty times. Further, *GBC* and *BNC* are both collections of edited texts, and so expected to be cleaner than the *GNC* which is a corpus extracted from the World Wide Web.

9 Conclusions and Future Work

We created a large word–colour association lexicon by crowdsourcing, which we will make publicly available. Word-choice questions were used to guide the annotators to the desired senses of the target words, and also as a gold questions for identifying malicious annotators (a common problem in crowdsourcing). We found that more than 32% of the words and 33% of the *Macquarie Thesaurus* categories have a strong association with one of the eleven colours chosen for the experiment. We analyzed abstract concepts, emotions in particular, and showed that they too have strong colour associations. Thus, using the right colours in tasks such as information visualization and web development, can not only improve semantic coherence but also inspire the desired emotional response.

Interestingly, we found that frequencies of colour associations follow the same order in which colour terms occur in different languages (Berlin and Kay, 1969). The frequency-based ranking of colour terms in the *BNC, GNC*, and *GBC* also had a high correlation with the Berlin and Kay order.

Finally, we show that automatic methods that rely on co-occurrence and polarity cues alone, and no labeled information of word–colour association, can accurately estimate the colour associated with a concept more than 60% of the time. The random choice and supervised baselines for this task are 9.1% and 33.3%, respectively. We are interested in using word–colour associations as a feature in sentiment analysis and for paraphrasing.

Acknowledgments

This research was funded by the National Research Council Canada. Grateful thanks to Peter Turney, Tara Small, and the reviewers for many wonderful ideas. Thanks to the thousands of people who answered the colour survey with diligence and care.

References

Satanjeev Banerjee and Ted Pedersen. 2003. Extended gloss overlaps as a measure of semantic relatedness. In *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-03)*, pages 805–810, Acapulco, Mexico.

Brent Berlin and Paul Kay. 1969. *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.

J.R.L. Bernard, editor. 1986. *The Macquarie Thesaurus*. Macquarie Library, Sydney, Australia.

Marc H. Bornstein. 1985. On the development of color naming in young children: Data and theory. *Brain and Language*, 26(1):72 – 93.

Roger W. Brown and Eric H. Lenneberg. 1954. A study in language and cognition. *Journal of Abnormal Psychology*, 49(3):454–462.

Lou Burnard. 2000. *Reference Guide for the British National Corpus (World Edition)*. Oxford University Computing Services.

Chris Callison-Burch. 2009. Fast, cheap and creative: Evaluating translation quality using amazon's mechanical turk. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, pages 286–295, Singapore.

Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, editors. 1999. *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers Inc., San Francisco, CA.

Wei-bin Chen. 2005. Comparative studies on cultural meaning difference of colors between china and western societies. *Journal of Fujian Institute of Socialism*.

Irvin L. Child, Jens A. Hansen, and Frederick W. Hornbeck. 1968. Age and sex differences in children's color preferences. *Child Development*, 39(1):237–247.

Richard E. Christ. 1975. Review and analysis of color coding research for visual displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 17:542–570.

Max Coltheart. 1981. The mrc psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33A:497–505.

John Gage. 1969. *Color and Culture: Practice and Meaning from Antiquity to Abstraction*. University of California Press, Ewing, NJ.

Jay J. Jiang and David W. Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of International Conference on Research on Computational Linguistics (ROCLING X)*, Taiwan.

Paul Kay and Luisa Maffi. 1999. Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist*, 101:743–760.

Naz Kaya. 2004. Relationship between color and emotion: a study of college students. *College Student Journal*, pages 396–405.

Dekang Lin. 1997. Using syntactic dependency as local context to resolve word sense ambiguity. In *Proceedings of the 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL/EACL-97)*, pages 64–71, Madrid, Spain.

Max Luscher. 1969. *The Luscher Color Test*. Random House, New York, New York.

Aaron Marcus. 1982. Color: a tool for computer graphics communication. *The Computer Image*, pages 76–90.

Barbara J. Meier. 1988. Ace: a color expert system for user interface design. In *Proceedings of the 1st annual ACM SIGGRAPH symposium on User Interface Software*, UIST '88, pages 117–128, New York, NY, USA. ACM.

Jean-Baptiste Michel, Yuan K. Shen, Aviva P. Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez L. Aiden. 2011. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014):176–182.

Saif Mohammad and Peter Turney. 2010. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, LA, California.

Saif Mohammad, Cody Dunne, and Bonnie Dorr. 2009. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP-2009)*, pages 599–608, Singapore.

Saif M. Mohammad. 2011. Even the abstract have colour: Consensus in wordcolour associations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, OR, USA.

Li-Chen Ou, M. Ronnier Luo, Pei-Li Sun, Neng-Chung Hu, and Hung-Shing Chen. 2011. Age effects on colour emotion, preference, and harmony. *Color Research and Application*, pages n/a–n/a.

Ted Pedersen, Siddharth Patwardhan, and Jason Michelizzi. 2004. WordNet::Similarity - Measuring the Relatedness of Concepts. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence*

(Intelligent Systems Demonstrations), pages 1024–1025, San Jose, CA, July.

Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. *Emotion: Theory, research, and experience*, 1(3):3–33.

Norma S. Pribadi, Maria G. Wadlow, and Daniel Boyarski. 1990. The use of color in computer interfaces: Preliminary research.

Carl Ratner. 1989. A sociohistorical critique of naturalistic theories of color perception. *Journal of Mind and Behavior*, 10(4):361–373.

Paul Sable and Okan Akcay. 2010. Color: Cross cultural marketing perspectves as to what governs our response to it. pages 950–954, Las vegas, CA.

Rion Snow, Brendan O'Connor, Daniel Jurafsky, and Andrew Ng. 2008. Cheap and fast - but is it good? Evaluating nonexpert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2008)*, pages 254–263, Waikiki, Hawaii.

Carlo Strapparava and Gozde Ozbal, 2010. *The Color of Emotions in Texts*, pages 28–32. Coling 2010 Organizing Committee.

J. H. Xin, K. M. Cheng, G. Taylor, T. Sato, and A. Hansuebsai. 2004. Cross-regional comparison of colour emotions part I: Quantitative analysis. *Color Research and Application*, 29(6):451–457.

David Yarowsky. 1992. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, pages 454–460, Nantes, France.