

Even the Abstract Have Colour: Consensus in Word-Colour Associations

Saif Mohammad, National Research Council Canada

Email: saif.mohammad@nrc-cnrc.gc.ca Web: <http://www.purl.org/net/saif.mohammad>

In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, June 2011.



1. Introduction

Many real-world concepts have associations with colour.

Concrete concepts:



→ white

iceberg



→ green

vegetation

Abstract concepts:



→ red

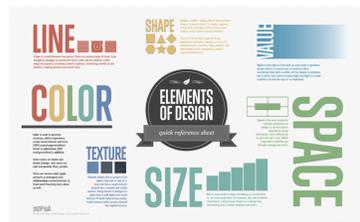
danger



→ white

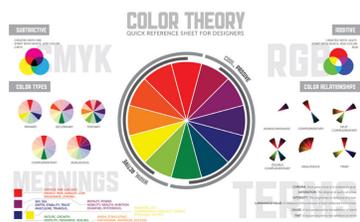
honesty

2. Motivation



Colour is a vital component of design in:

- information visualization (Card et al., 1999)
- product marketing (Sable and Akcay, 2010)
- webpage design (Pribadi et al., 1990)



Colours complement linguistic information:

- strengthen the message (improve semantic coherence)
- ease cognitive load on the receiver
- convey the message quickly
- evoke the desired emotional response



Examples of colour terms in language given the **red** carpet (given special treatment) through **rose-tinted** glasses (being optimistic) turned **green** with envy (was envious) **grey** with uncertainty (uncertain) [from Bianca Madison's poem *Confusion*]

Word-colour associations can help:

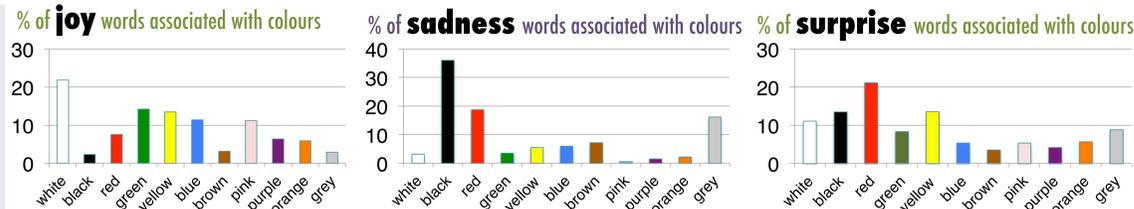
- textual entailment
- paraphrasing
- machine translation
- sentiment analysis
- emotion detection

6. Colour Signature of Emotions

Combined the term-colour lexicon with the NRC term-emotion lexicon (Mohammad and Turney, 2011).

NRC Emotion Lexicon:

- eight basic emotions
- two polarities (pos, neg)
- 24,200 word-sense pairs
- 14,200 word types



3. Crowdsourcing a Word-Colour Association Lexicon

Mechanical Turk questionnaire:

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

- car - tree - nap - wave

Q2. Which colour is associated with *sleep*?

- black - green - purple... (in random order)

Options: study on colour in language (Berlin and Kay, 1969)

- 1. white, 2. black, 3. red, 4. green, 5. yellow, 6. blue, 7. brown, 8. pink, 9. purple, 10. orange, 11. grey
- no "not associated with any colour" option

Gold question: generated using a thesaurus

- guides Turkers to desired sense
- aides quality control
- if Q1 is answered incorrectly (10% cases):
 - response to Q2 is discarded
 - Turker is immediately notified

Target word:

- must occur in the thesaurus, and
- must occur in the *Google n-gram corpus* (frequency > 120,000)

4. NRC Word-Colour Association Lexicon

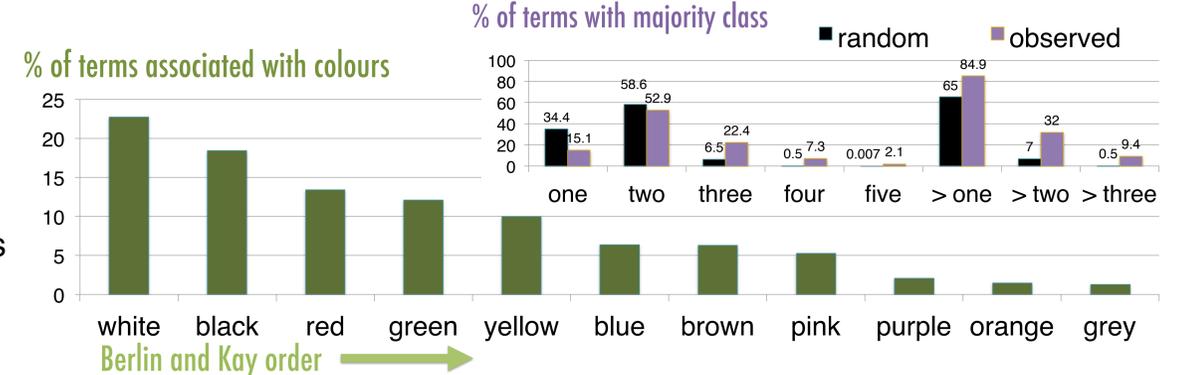
Version 0.1: *Macquarie Thesaurus* terms

- 8,800 word-sense pairs
- 4,500 word types
- 4.45 annotations/term

Version 0.2: *Roget's Thesaurus* terms

- 24,200 word-sense pairs
- 14,200 word types
- 5.1 annotations/term

Majority class: 1 (max disagreement), 2, 3, 4, 5 (max agreement)



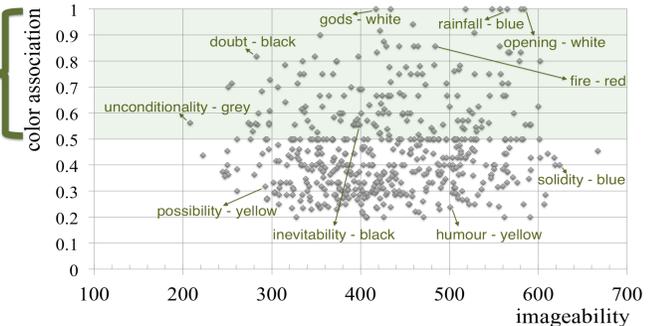
For corpus-based automatic methods: "Colourful Language: Measuring Word-Colour Associations" Saif Mohammad, In Proceedings of the ACL 2011 Workshop on Cognitive Modeling and Computational Linguistics (CMCL).

5. Imageability and Colour Association

Colour association of a thesaurus category:

- proportion of words associated with dominant colour
- imageability of a category: average of constituent of words
- MRCP database (Coltheart, 1981): ratings of 9240 words
- scale: 100 (hard to visualize) to 700 (easy to visualize)

Scatter plot of thesaurus categories. $r = 0.116$.



7. Conclusions

- created word-colour lexicon by crowdsourcing
- about 32% of the words, and 33% of thesaurus categories had strong colour associations
- abstract concepts also have associations
- frequencies follow the Berlin and Kay order

8. Future Work

Show usefulness in:

- NLP tasks
- sentiment analysis
- textual entailment
- Information visualization