Colourful Language: Measuring Word-Colour Associations

Saif Mohammad
National Research Council Canada
Examples of Concrete Concepts

- **iceberg**
- **vegetation**

- white
- green
Examples of Abstract Concepts

- red
- white
- danger
- honesty
Road Map

- Introduction and Motivation
- Related Work
- Manual Annotation
  - Analysis and findings
- Manifestation of associations in WordNet and in text
  - Automatic methods
Good Design

Colour is a vital component of:

- information visualization (Christ, 1975; Card et al., 1999)
- product marketing (Sable and Akcay, 2010)
- webpage design (Meier, 1988; Pribadi et al., 1990)

“It’s always good to be able to articulate design choices to your clients; why you put something where, why you chose the color scheme you did, etc. This is one of the biggest differences between a designer and a non-designer.”

-- Jeff Archibald
(founder of Paper Leaf, a graphic- and web-design company)
Source: Paper Leaf
Colour Choices

Source: Paper Leaf
Colours can Complement Linguistic Information

- Strengthens the message (improves semantic coherence)
- Eases cognitive load on the receiver
- Conveys the message quickly
- Evokes the desired emotional response
Expressions Involving Colour

*turned green with envy* (was envious)
*given the red carpet* (given special treatment)
*looking through rose-tinted glasses* (being optimistic)
*grey with uncertainty* (uncertain)
[from Bianca Madison’s poem *Confusion*]

Concept–colour associations may also help:
- textual entailment
- paraphrasing
- machine translation
- sentiment analysis
Related Work

- On word-colour associations:
  - Academic: nothing on a large scale
  - Commercial: Cymbolism

- On colour, language, and cognition:
  Brown and Lenneberg, 1954; Ratner, 1989; Bornstein, 1985

- On age and gender preferences for colour:
  Child et al. 1968; Ou et al. 2011

- On emotions evoked by colour:
  Luscher, 1969; Xin et al., 2004; Kaya, 2004
Related Work (continued)

  - If a language has only two colours: white and black.
  - If a language has three: white, black, red.
  - And so on till eleven colours.

- Berlin and Kay order:
  1. white, 2. black, 3. red, 4. green, 5. yellow, 6. blue,
  7. brown, 8. pink, 9. purple, 10. orange, 11. grey

- We used these eleven colours in our annotations.
  - Hundreds more:
<table>
<thead>
<tr>
<th>Name</th>
<th>Hex triplet</th>
<th>Red</th>
<th>Green</th>
<th>Blue</th>
<th>Hue</th>
<th>Satur</th>
<th>Light</th>
<th>Satur</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Force blue</td>
<td>#5D8AA8</td>
<td>36%</td>
<td>54%</td>
<td>66%</td>
<td>204°</td>
<td>30%</td>
<td>51%</td>
<td>45%</td>
<td>66%</td>
</tr>
<tr>
<td>Alice blue</td>
<td>#00FFFF</td>
<td>94%</td>
<td>97%</td>
<td>100%</td>
<td>208°</td>
<td>100%</td>
<td>97%</td>
<td>6%</td>
<td>100%</td>
</tr>
<tr>
<td>Alizarin crimson</td>
<td>#E32636</td>
<td>82%</td>
<td>10%</td>
<td>26%</td>
<td>231°</td>
<td>78%</td>
<td>46%</td>
<td>187%</td>
<td>110%</td>
</tr>
<tr>
<td>Almond</td>
<td>#EEDECD</td>
<td>94%</td>
<td>87%</td>
<td>80%</td>
<td>30°</td>
<td>52%</td>
<td>87%</td>
<td>14%</td>
<td>94%</td>
</tr>
<tr>
<td>Amaranth</td>
<td>#E52B50</td>
<td>90%</td>
<td>17%</td>
<td>31%</td>
<td>348°</td>
<td>78%</td>
<td>53%</td>
<td>81%</td>
<td>90%</td>
</tr>
<tr>
<td>Amber</td>
<td>#FFB600</td>
<td>100%</td>
<td>75%</td>
<td>0%</td>
<td>45°</td>
<td>100%</td>
<td>50%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Amber (SAE/ECE)</td>
<td>#FF7E00</td>
<td>100%</td>
<td>49%</td>
<td>0%</td>
<td>30°</td>
<td>100%</td>
<td>50%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>American rose</td>
<td>#FF0033</td>
<td>100%</td>
<td>1%</td>
<td>24%</td>
<td>345°</td>
<td>100%</td>
<td>51%</td>
<td>99%</td>
<td>87%</td>
</tr>
<tr>
<td>Amethyst</td>
<td>#9966CC</td>
<td>60%</td>
<td>40%</td>
<td>80%</td>
<td>270°</td>
<td>50%</td>
<td>60%</td>
<td>50%</td>
<td>80%</td>
</tr>
<tr>
<td>Android Green</td>
<td>#A4C639</td>
<td>64%</td>
<td>78%</td>
<td>22%</td>
<td>74°</td>
<td>55%</td>
<td>50%</td>
<td>7%</td>
<td>78%</td>
</tr>
<tr>
<td>Anti-flash white</td>
<td>#F2F3F4</td>
<td>95%</td>
<td>95%</td>
<td>96%</td>
<td>210°</td>
<td>8%</td>
<td>95%</td>
<td>1%</td>
<td>96%</td>
</tr>
<tr>
<td>Antique brass</td>
<td>#CD9575</td>
<td>80%</td>
<td>58%</td>
<td>46%</td>
<td>22°</td>
<td>47%</td>
<td>63%</td>
<td>43%</td>
<td>80%</td>
</tr>
<tr>
<td>Antique fuchsia</td>
<td>#915C83</td>
<td>57%</td>
<td>36%</td>
<td>51%</td>
<td>316°</td>
<td>22%</td>
<td>47%</td>
<td>37%</td>
<td>57%</td>
</tr>
<tr>
<td>Antique white</td>
<td>#FAEBD7</td>
<td>98%</td>
<td>92%</td>
<td>84%</td>
<td>34°</td>
<td>78%</td>
<td>91%</td>
<td>14%</td>
<td>98%</td>
</tr>
<tr>
<td>Ao (English)</td>
<td>#000000</td>
<td>0%</td>
<td>50%</td>
<td>0%</td>
<td>120°</td>
<td>100%</td>
<td>25%</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>Apple green</td>
<td>#80B600</td>
<td>55%</td>
<td>71%</td>
<td>0%</td>
<td>74°</td>
<td>100%</td>
<td>36%</td>
<td>100%</td>
<td>50%</td>
</tr>
<tr>
<td>Apricot</td>
<td>#FBCEB1</td>
<td>98%</td>
<td>81%</td>
<td>69%</td>
<td>24°</td>
<td>90%</td>
<td>84%</td>
<td>29%</td>
<td>98%</td>
</tr>
<tr>
<td>Aqua</td>
<td>#00FFFF</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
<td>180°</td>
<td>100%</td>
<td>50%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Aquamarine</td>
<td>#7FFFD0</td>
<td>50%</td>
<td>100%</td>
<td>83%</td>
<td>160°</td>
<td>100%</td>
<td>75%</td>
<td>50%</td>
<td>100%</td>
</tr>
<tr>
<td>Army green</td>
<td>#4B5320</td>
<td>29%</td>
<td>33%</td>
<td>13%</td>
<td>69°</td>
<td>44%</td>
<td>23%</td>
<td>61%</td>
<td>33%</td>
</tr>
<tr>
<td>Arsenic</td>
<td>#38444B</td>
<td>23%</td>
<td>27%</td>
<td>29%</td>
<td>206°</td>
<td>12%</td>
<td>26%</td>
<td>21%</td>
<td>29%</td>
</tr>
<tr>
<td>Arylide yellow</td>
<td>#E9D66B</td>
<td>91%</td>
<td>84%</td>
<td>42%</td>
<td>51°</td>
<td>74%</td>
<td>67%</td>
<td>54%</td>
<td>91%</td>
</tr>
<tr>
<td>Ash grey</td>
<td>#B2EB55</td>
<td>70%</td>
<td>75%</td>
<td>71%</td>
<td>135°</td>
<td>9%</td>
<td>72%</td>
<td>6%</td>
<td>75%</td>
</tr>
<tr>
<td>Asparagus</td>
<td>#87A96B</td>
<td>53%</td>
<td>66%</td>
<td>42%</td>
<td>93°</td>
<td>27%</td>
<td>54%</td>
<td>37%</td>
<td>66%</td>
</tr>
<tr>
<td>Atomic tangerine</td>
<td>#FF9966</td>
<td>100%</td>
<td>60%</td>
<td>40%</td>
<td>20°</td>
<td>100%</td>
<td>70%</td>
<td>60%</td>
<td>100%</td>
</tr>
<tr>
<td>Auburn</td>
<td>#6D351A</td>
<td>43%</td>
<td>21%</td>
<td>10%</td>
<td>20°</td>
<td>62%</td>
<td>27%</td>
<td>76%</td>
<td>43%</td>
</tr>
<tr>
<td>Aureolín</td>
<td>#FDEE00</td>
<td>99%</td>
<td>93%</td>
<td>0%</td>
<td>56°</td>
<td>100%</td>
<td>50%</td>
<td>100%</td>
<td>99%</td>
</tr>
</tbody>
</table>
Manual Annotation and Analysis
Crowdsourcing

- Annotations:
  Amazon’s Mechanical Turk: 5 annotations per term
- Target terms:
  Macquarie Thesaurus, Google N-gram Corpus
- Questionnaire:
  Q1. Which word is closest in meaning to sleep?
  - car  - tree  - nap  - king
  Q2. Which colour is associated with sleep?
  - black  - green  - purple…
  … (11 colour options in random order)
  - No “not associated with any colour” option.
Post-processing

- Annotations discarded due to Q1:
  - about 10%
- Other discards:
  - terms with less than 3 valid annotations
- Remaining set:
  - annotations for 8,813 word-sense pairs
- Valid annotations per term:
  - 4.45
Associations with Colours

% of annotations

% of terms

Berlin and Kay order
Agreement

- Majority class:
  1 (maximum disagreement), 2, 3, 4, 5 (maximum agreement)
- Random annotation and observed percentages of the majority class:

% of terms

- % of terms:
  - one: 15.1%
  - two: 58.6%
  - three: 22.4%
  - four: 0.5%
  - five: 0.007%
  - > one: 65%
  - > two: 32%
  - > three: 0.5%

Colourful Language. Saif Mohammad.
Thesaurus Categories

- Sets of closely related words
- For each category
  - determined the colour $c$ most associated with it
- Strength of color association of a category $cat$:
  $$\text{# of words in } cat \text{ associated with } c \over \text{# of words in the } cat$$
- 33.1% of the *Macquarie Thesaurus* categories had an association greater than 0.5
  - Gold standard category-colour associations
Imageability and Colour Association

Is there a correlation between imageability and tendency to have a colour association?

- MRC Psycholinguistic Database (Coltheart, 1981)
  - imageability ratings: 9240 words
  - scale: 100 (hard to visualize) to 700 (easy to visualize)

- Imageability of a thesaurus category:
  - Average imageability of its constituent words
Scatter Plot of Thesaurus Categories

Pearson’s product moment correlation: 0.116
Do emotion words have a colour association?

- Combined the term-colour lexicon with the term-emotion lexicon (Mohammad and Turney, 2010)
- Determined the colours associated with emotion words.

% of surprise words associated with different colours
% of **joy** words associated with different colours

% of **sadness** words associated with different colours
% of **negative** words associated with different colours

% of **positive** words associated with different colours

Colourful Language. Saif Mohammad. 23
Manifestation of Word–Colour Associations in WordNet and in Text
Colours in WordNet

Are words and their associated colours close to each other in WordNet?

- *darkness*: hypernym of black
- *inflammation*: one hop away from red
WordNet-based Automatic Method

- Determine colour closest to target terms in WordNet
- Choose colour closest to most terms in a thesaurus category
- Compare with gold standard category-colour associations

Accuracy, in %

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>random</td>
<td></td>
</tr>
<tr>
<td>most associated</td>
<td>40</td>
</tr>
<tr>
<td>Jiang Conrath</td>
<td></td>
</tr>
<tr>
<td>Lin</td>
<td></td>
</tr>
<tr>
<td>Lesk</td>
<td></td>
</tr>
<tr>
<td>gloss vector</td>
<td>30</td>
</tr>
</tbody>
</table>

unsupervised baseline  supervised baseline  similarity measures  relatedness measures
Rank correlation with Berlin and Kay order:
Google N-gram Corpus (GNC): 0.884
Google Books Corpus (GBC): 0.918

Do words co-occur with their associated colours more often than any other colour?

- *darkness* with black
- *inflammation* with red
Corpus-based Automatic Method

- Determine colour that co-occurs most with target terms
  - Conditional probability
- Choose colour associated most with terms in a thesaurus category
- Compare with gold standard category-colour associations
Results

Accuracy, in %

- Above baselines, but not by that much.
- Can polarity help?

Colourful Language. Saif Mohammad.
% of negative words associated with different colours

% of positive words associated with different colours
Polarity Cues

- Updated algorithm:
  - If a term is positive:
    - co-occurrence is used to choose from only the positive colours
  - If a term is negative:
    - co-occurrence is used to choose from only the negative colours

- Macquarie Semantic Orientation Lexicon (MSOL) (Mohammad et al. 2009):
  - Automatically created
  - 76,400 terms marked as positive or negative
Results

Accuracy, in %

- unsupervised baselines
- supervised baseline
- WordNet-based methods
- corpus-based methods

with polarity cues

Colourful Language. Saif Mohammad.
Conclusions

- Created a large word-colour association lexicon by crowdsourcing
- More than 32% of the words, and 33% of thesaurus categories had strong colour associations
- Abstract concepts just as likely to have colour associations
- Frequencies of associations follow the Berlin and Kay order
  - As do frequencies of colour terms in corpora
- Automatic methods of association obtain 60% accuracy
  - Features: co-occurrence and polarity
  - Supervised baseline: 33.3%
Ongoing and Future Work

- Created a much larger lexicon
  - Source: Roget Thesaurus
  - Size: 24,000 word-sense pairs

- Improve performance of automatic methods
  - Other features? Image data?
  - Determine performance at word-level

- Show usefulness in NLP tasks
  - Sentiment analysis
  - Textual entailment
Ongoing and Future Work (continued)

- Consider theoretical questions
  - What do these analyses tell us about how we think about colour?
  - What do gender and age differences tell us? *(Child et al. 1968, Ou et al. 2011)*

- Release data for users at large
  - Information Visualization groups
  - Graphic- and web-design teams
  - Psychologists
  - Cognitive scientists