



Colourful Language:

Measuring Word-Colour Associations

Saif Mohammad

National Research Council Canada



Examples of Concrete Concepts



white

iceberg



green

vegetation

Examples of Abstract Concepts



danger

→ red



honesty

→ white



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:

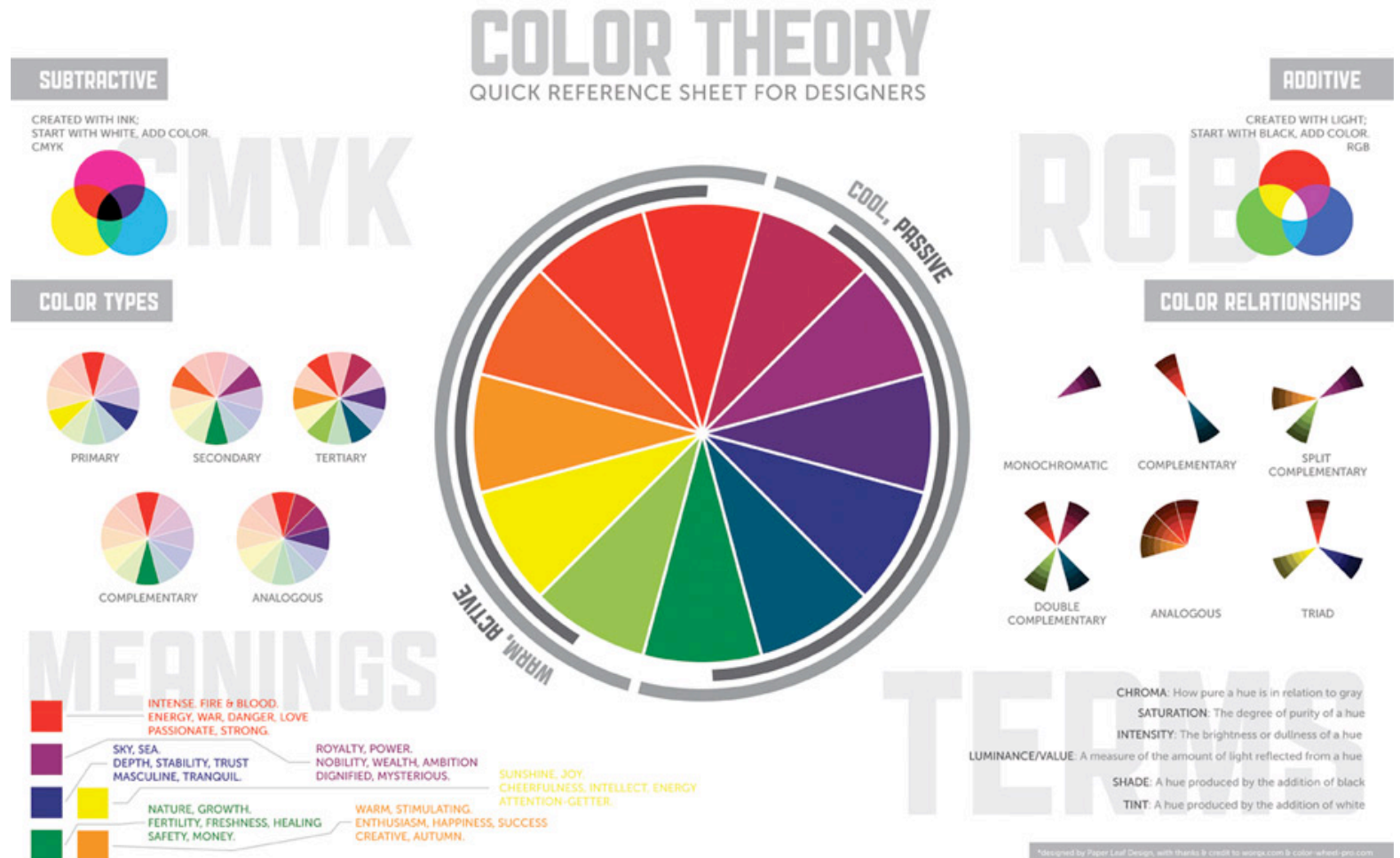
- information visualization (Christ, 1975; Card et al., 1999)
- product marketing (Sable and Akcay, 2010)
- webpage design (Meier, 1988; Pribadi et al., 1990)

“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*

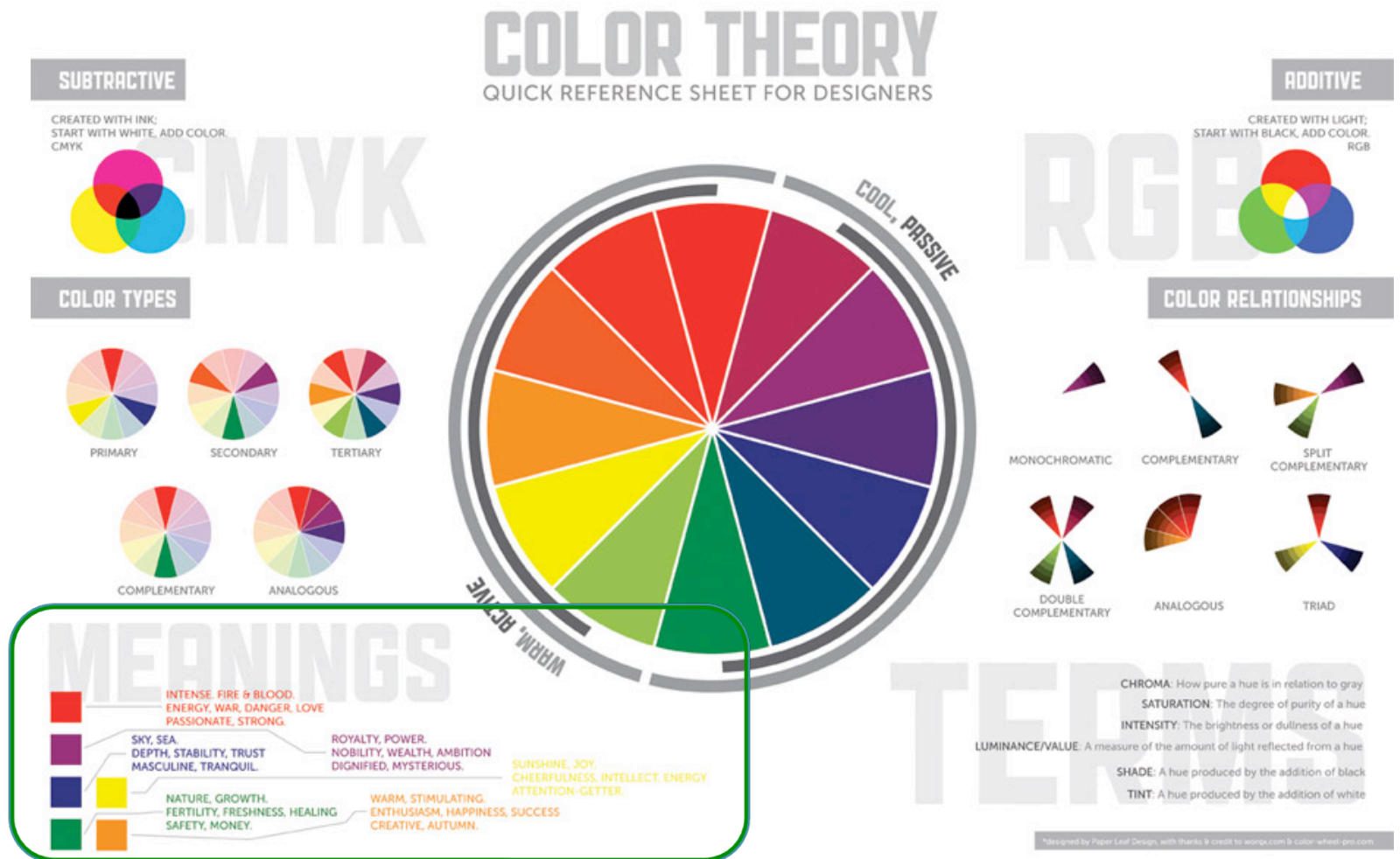
(founder of Paper Leaf, a graphic- and web-design company)

Colour Choices



Source: Paper Leaf

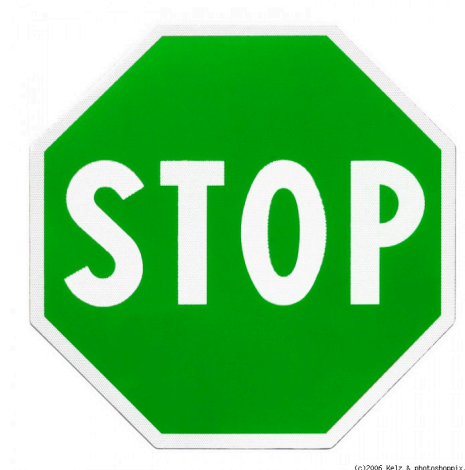
Colour Choices



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





























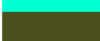
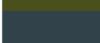







- On word-colour associations:
 - Academic: nothing on a large scale
 - Commercial: Symbolism
- On colour, language, and cognition:
Brown and Lenneberg, 1954; Ratner, 1989; Bornstein, 1985
- On age and gender preferences for colour:
Child et al. 1968; Ou et al. 2011
- On emotions evoked by colour:
Luscher, 1969; Xin et al., 2004; Kaya, 2004

Related Work (continued)

- Berlin and Kay, 1969, and later Kay and Maffi (1999)
 - 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:
http://en.wikipedia.org/wiki/List_of_colors

Just the A's

Color names

	Name 	Hex triplet 	Red 	Green 	Blue 	Hue 	Satur 	Light 	Satur 	Value 
	Air Force blue	#5D8AA8	36%	54%	66%	204°	30%	51%	45%	66%
	Alice blue	#F0F8FF	94%	97%	100%	208°	100%	97%	6%	100%
	Alizarin crimson	#E32636	82%	10%	26%	231°	78%	46%	187%	110%
	Almond	#EFDECD	94%	87%	80%	30°	52%	87%	14%	94%
	Amaranth	#E52B50	90%	17%	31%	348°	78%	53%	81%	90%
	Amber	#FFBF00	100%	75%	0%	45°	100%	50%	100%	100%
	Amber (SAE/ECE)	#FF7E00	100%	49%	0%	30°	100%	50%	100%	100%
	American rose	#FF033E	100%	1%	24%	345°	100%	51%	99%	87%
	Amethyst	#9966CC	60%	40%	80%	270°	50%	60%	50%	80%
	Android Green	#A4C639	64%	78%	22%	74°	55%	50%	7%	78%
	Anti-flash white	#F2F3F4	95%	95%	96%	210°	8%	95%	1%	96%
	Antique brass	#CD9575	80%	58%	46%	22°	47%	63%	43%	80%
	Antique fuchsia	#915C83	57%	36%	51%	316°	22%	47%	37%	57%
	Antique white	#FAEBD7	98%	92%	84%	34°	78%	91%	14%	98%
	Ao (English)	#008000	0%	50%	0%	120°	100%	25%	100%	50%
	Apple green	#8DB600	55%	71%	0%	74°	100%	36%	100%	50%
	Apricot	#FBCB11	98%	81%	69%	24°	90%	84%	29%	98%
	Aqua	#00FFFF	0%	100%	100%	180°	100%	50%	100%	100%
	Aquamarine	#7FFFD0	50%	100%	83%	160°	100%	75%	50%	100%
	Army green	#4B5320	29%	33%	13%	69°	44%	23%	61%	33%
	Arsenic	#3B444B	23%	27%	29%	206°	12%	26%	21%	29%
	Arylide yellow	#E9D66B	91%	84%	42%	51°	74%	67%	54%	91%
	Ash grey	#B2BEB5	70%	75%	71%	135°	9%	72%	6%	75%
	Asparagus	#87A96B	53%	66%	42%	93°	27%	54%	37%	66%
	Atomic tangerine	#FF9966	100%	60%	40%	20°	100%	70%	60%	100%
	Auburn	#6D351A	43%	21%	10%	20°	62%	27%	76%	43%
	Aureolin	#FDEE00	99%	93%	0%	56°	100%	50%	100%	99%



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...

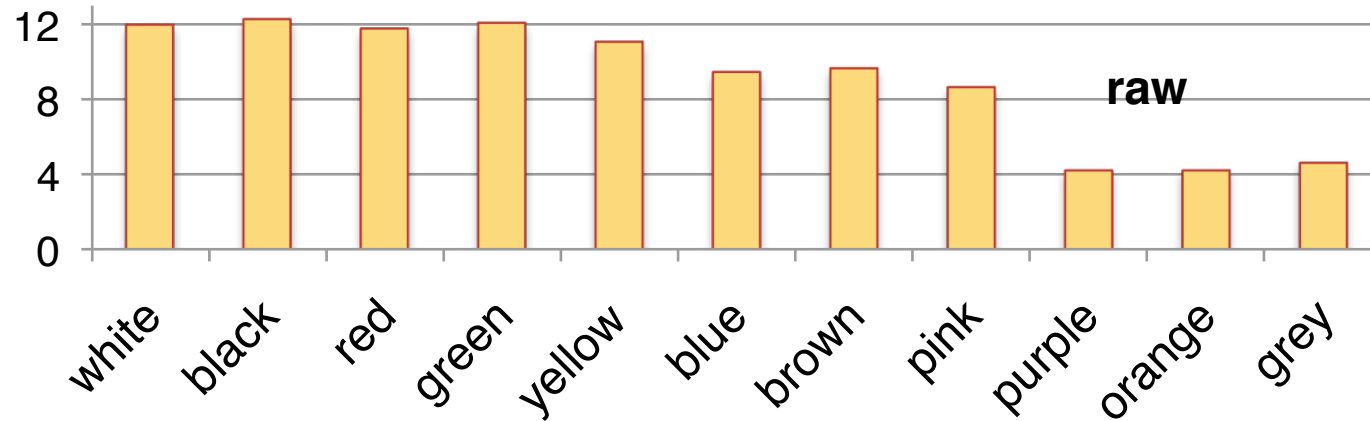
... (11 colour options in random order)
- No “not associated with any colour” option.

Post-processing

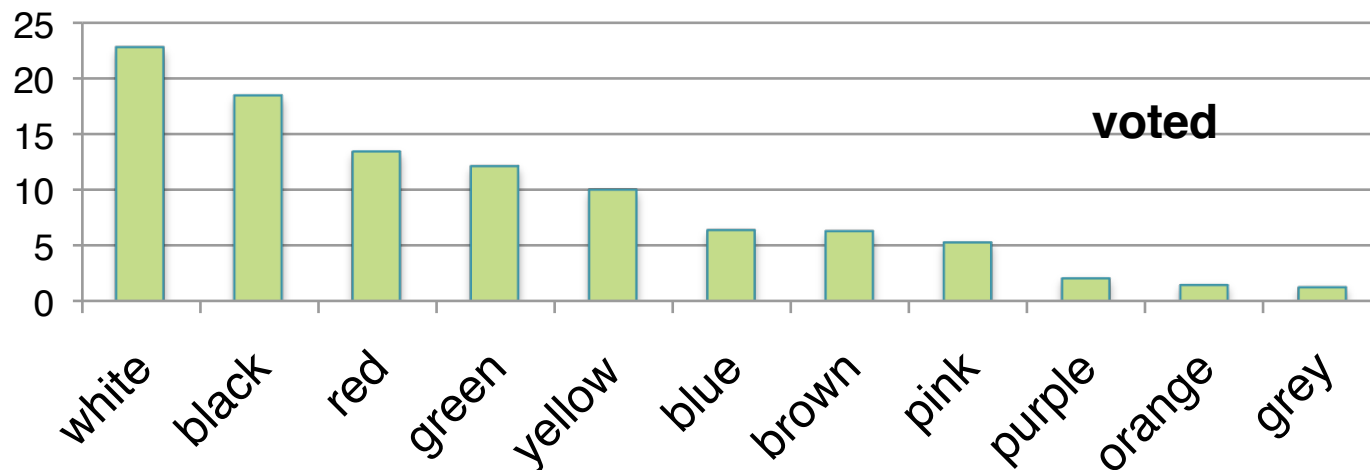
- Annotations discarded due to Q1:
 - about 10%
- Other discards:
 - terms with less than 3 valid annotations
- Remaining set:
 - annotations for 8,813 word-sense pairs
- Valid annotations per term:
 - 4.45

Associations with Colours

% of annotations



% of terms



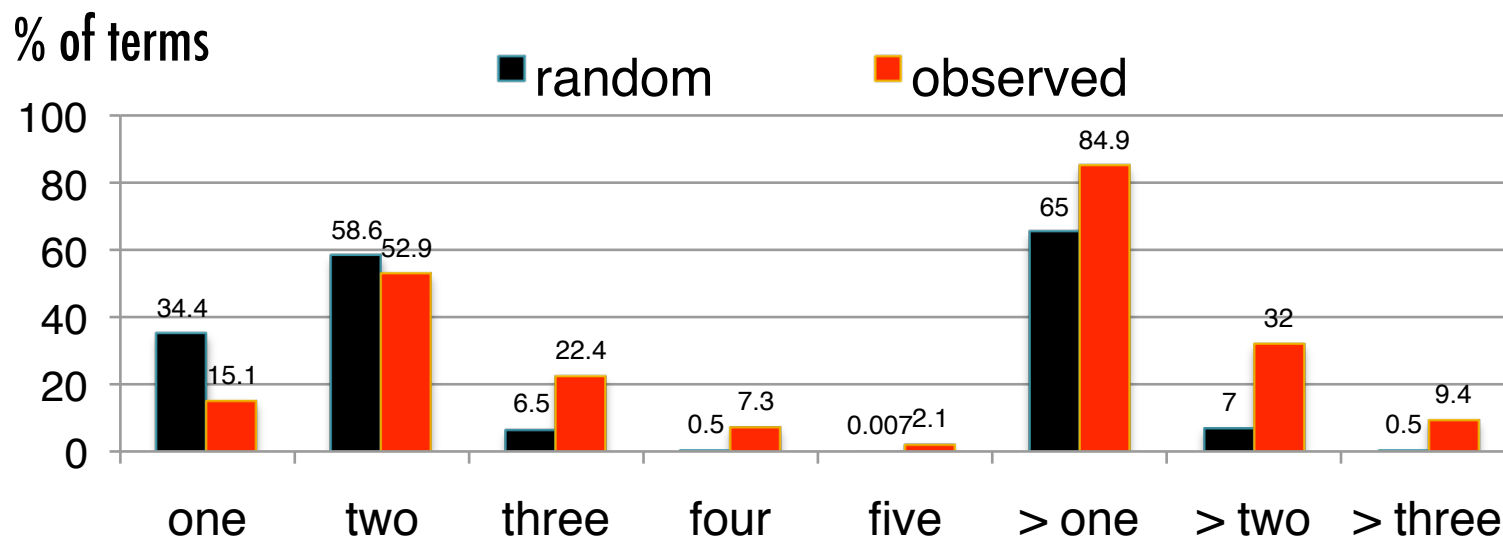
Berlin and Kay order



Colourful Language. Saif Mohammad.

Agreement

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



Thesaurus Categories

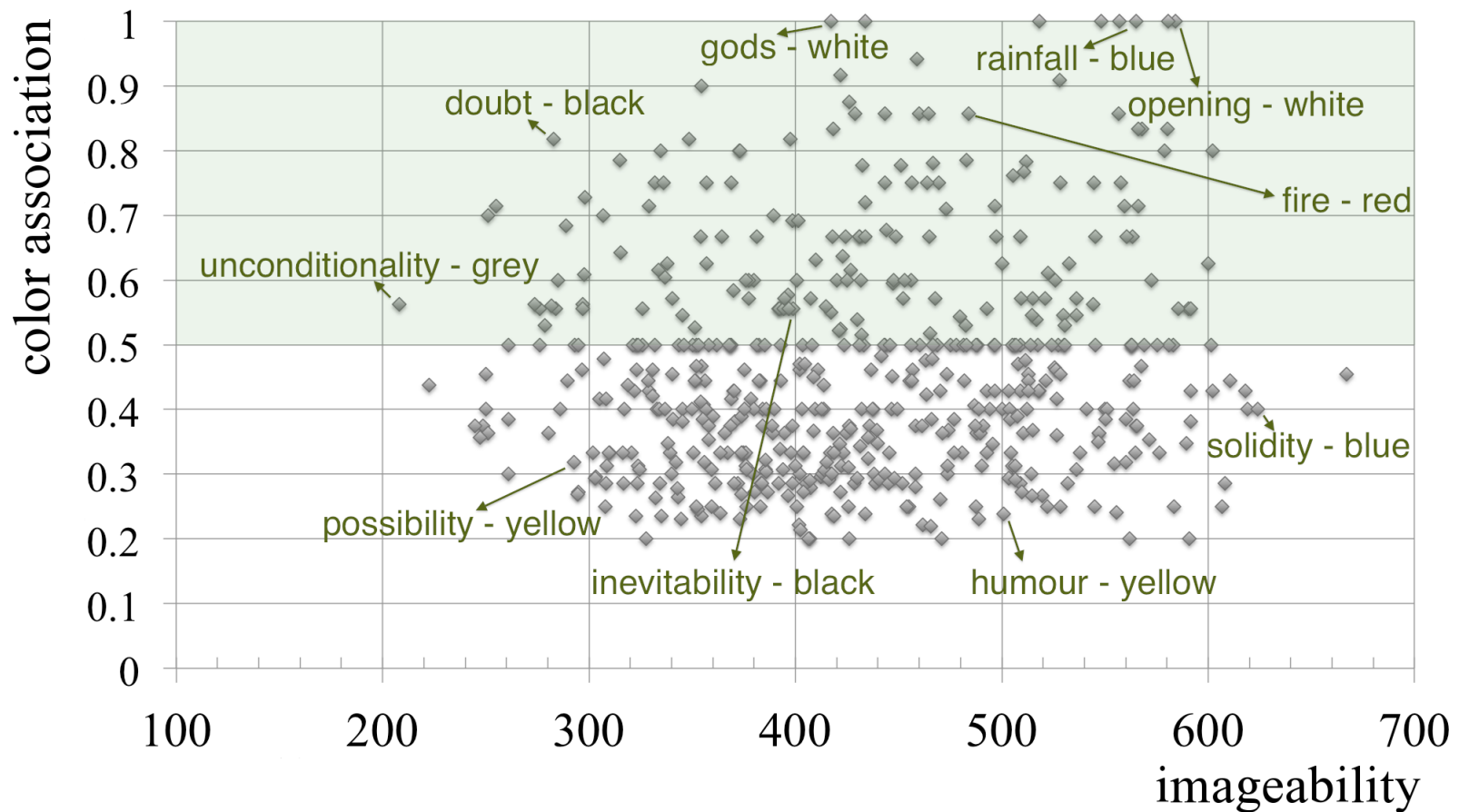
- Sets of closely related words
- For each category
 - determined the colour *c* most associated with it
- Strength of color association of a category *cat*:
$$= \frac{\text{\# of words in } cat \text{ associated with } c}{\text{\# of words in the } cat}$$
- 33.1% of the *Macquarie Thesaurus* categories 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

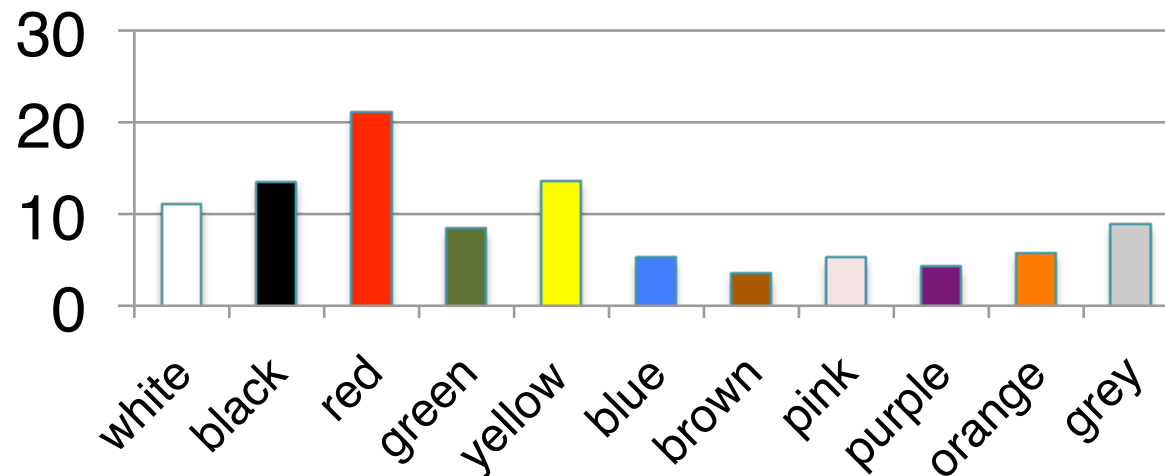


Pearson's product moment correlation: 0.116

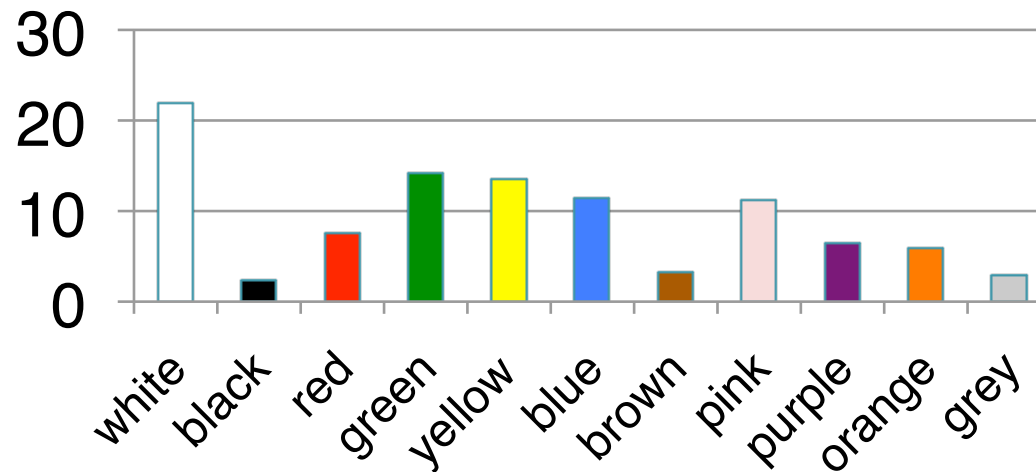
Do emotion words have a colour association?

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

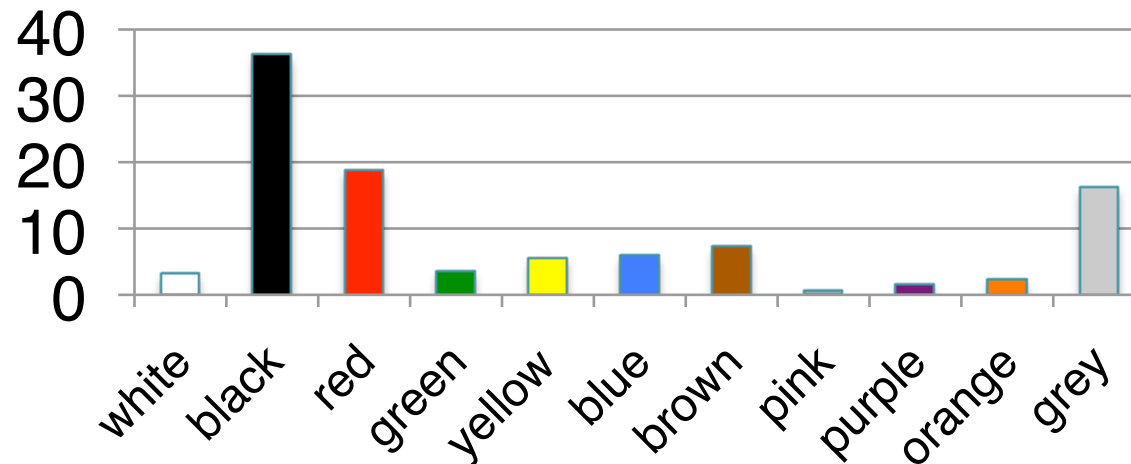
% of **surprise** words associated with different colours



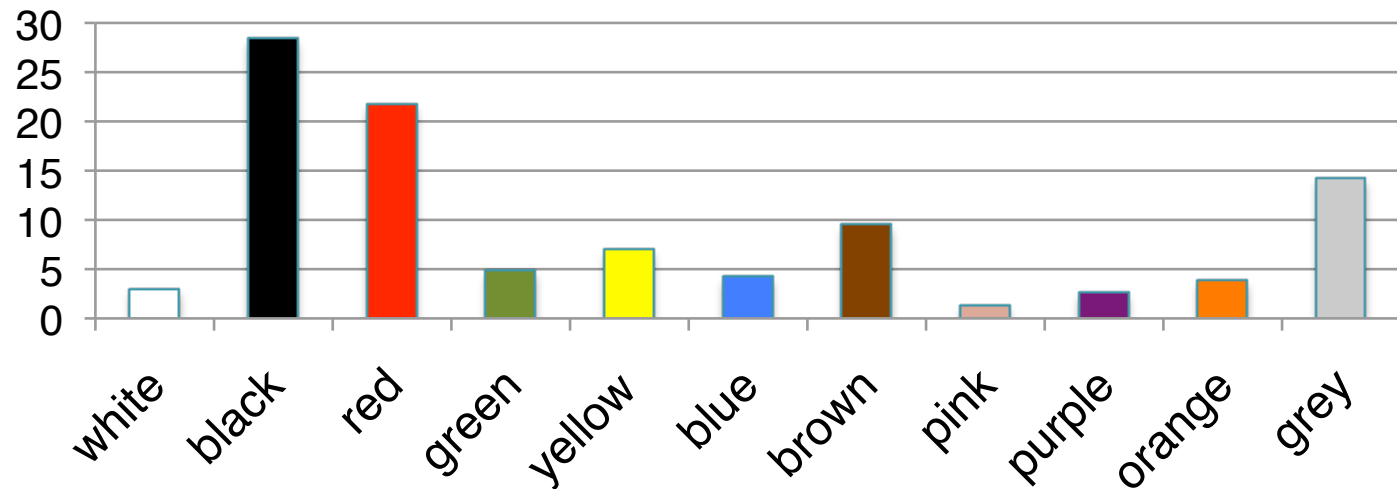
% of **joy** words associated with different colours



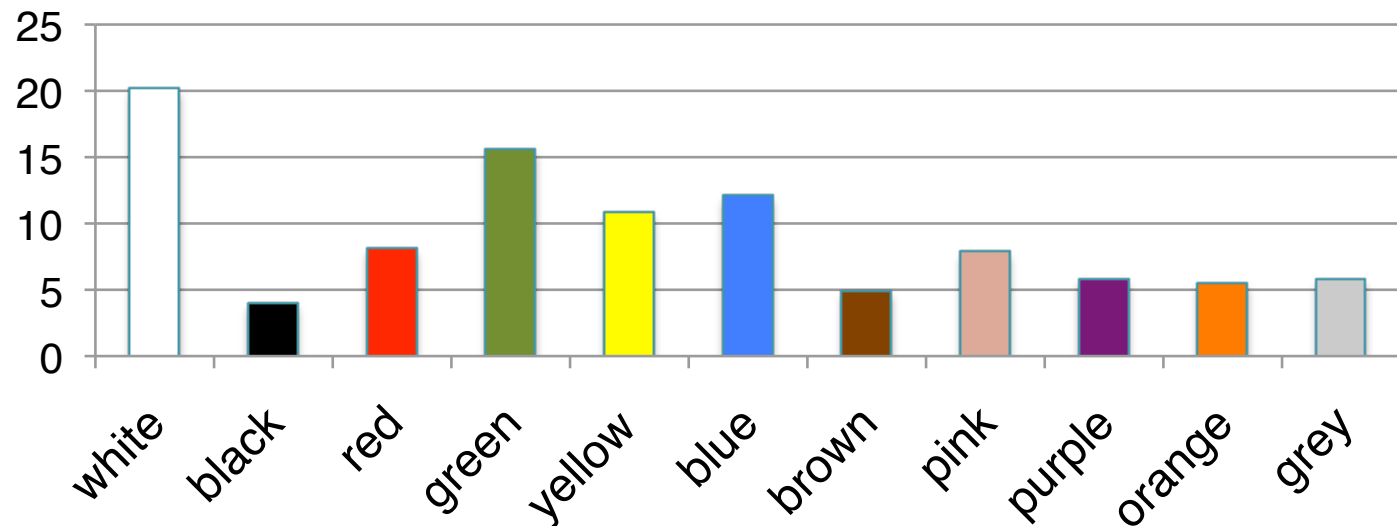
% of **sadness** words associated with different colours



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



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

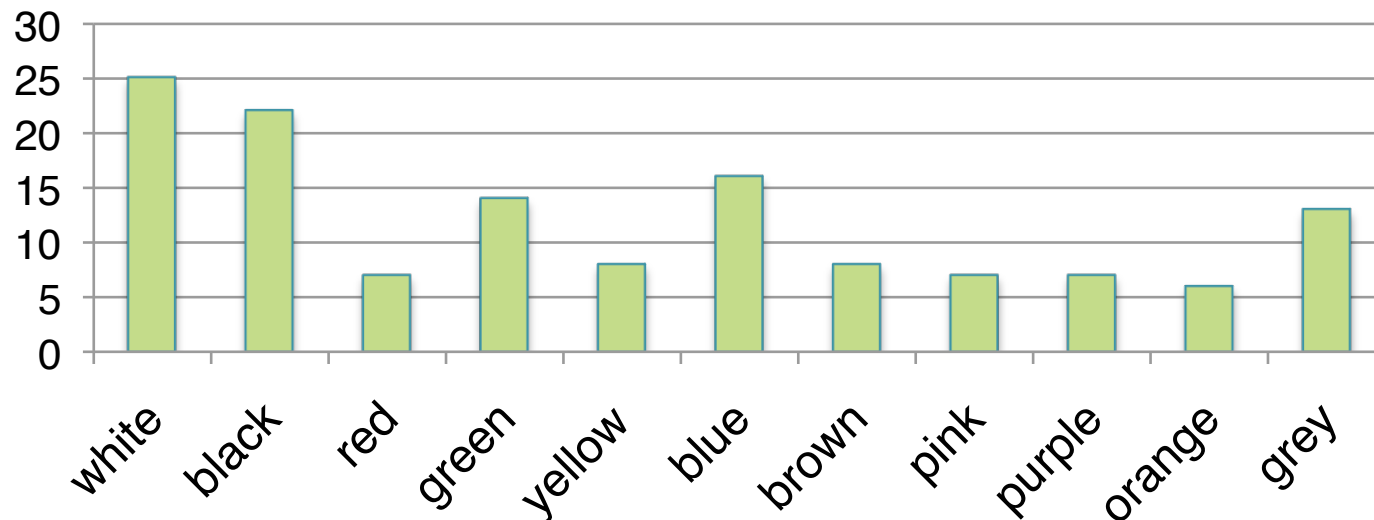




Manifestation of Word–Colour Associations in WordNet and in Text

Colours in WordNet

of senses



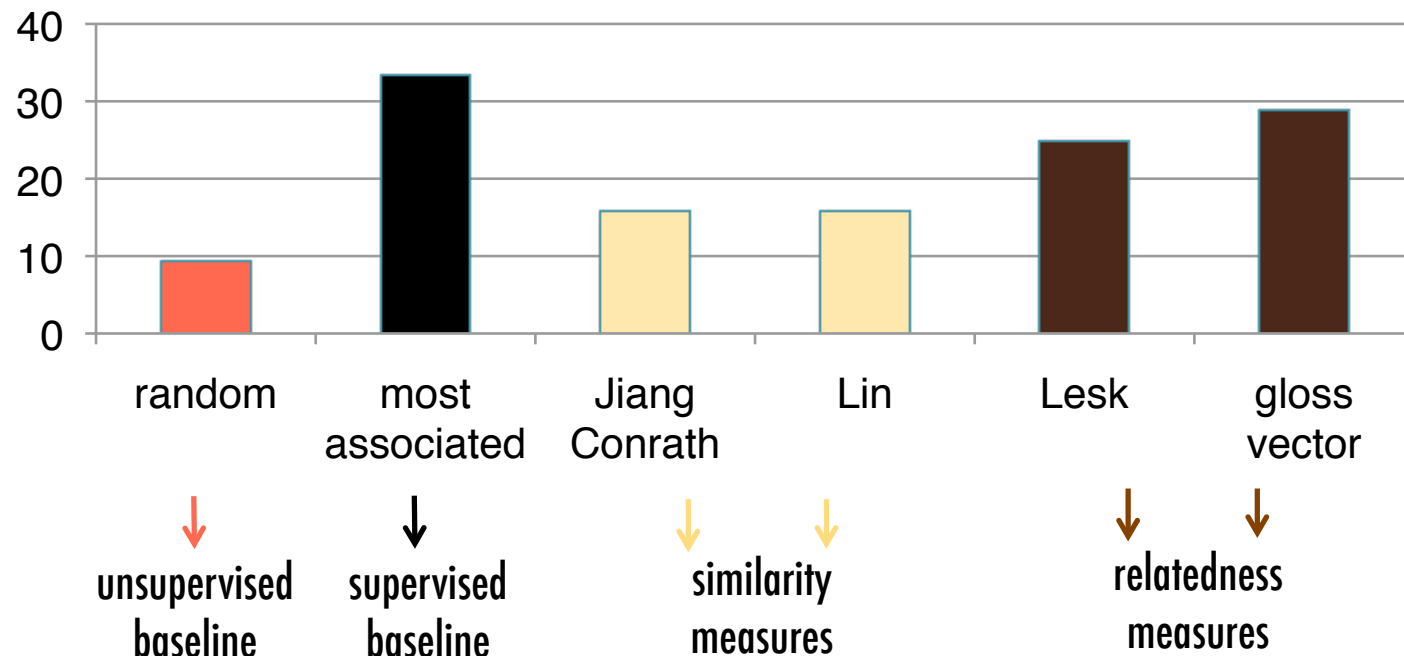
Are words and their associated colours close to each other in WordNet?

- *darkness*: hypernym of black
- *inflammation*: one hop away from red

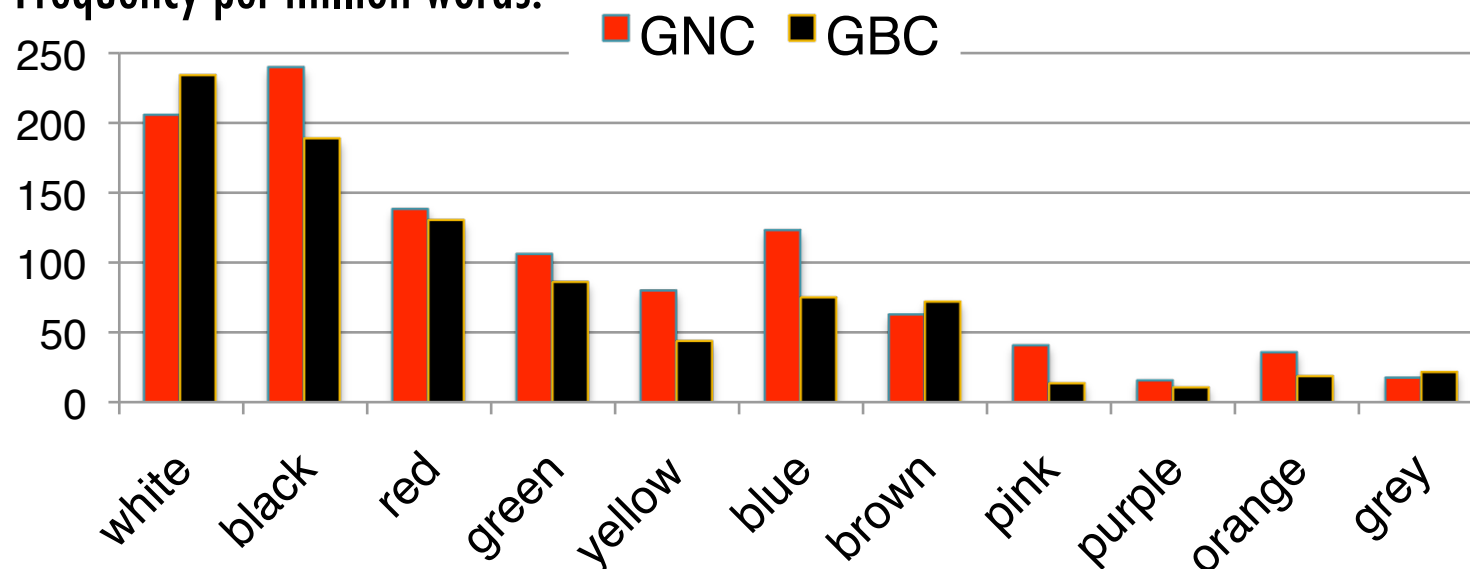
WordNet-based Automatic Method

- 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 %



Frequency per million words.



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

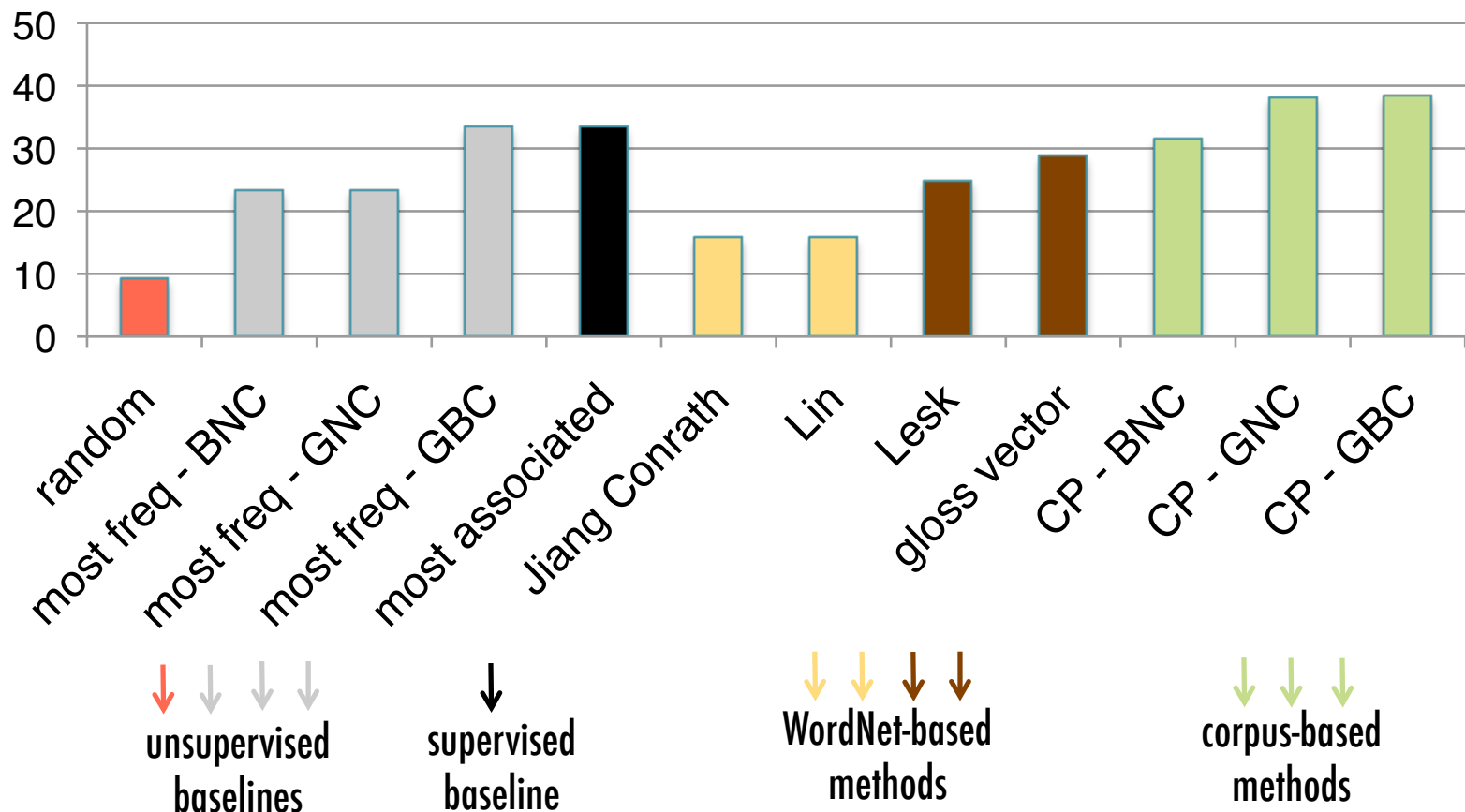


Corpus-based Automatic Method

- 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

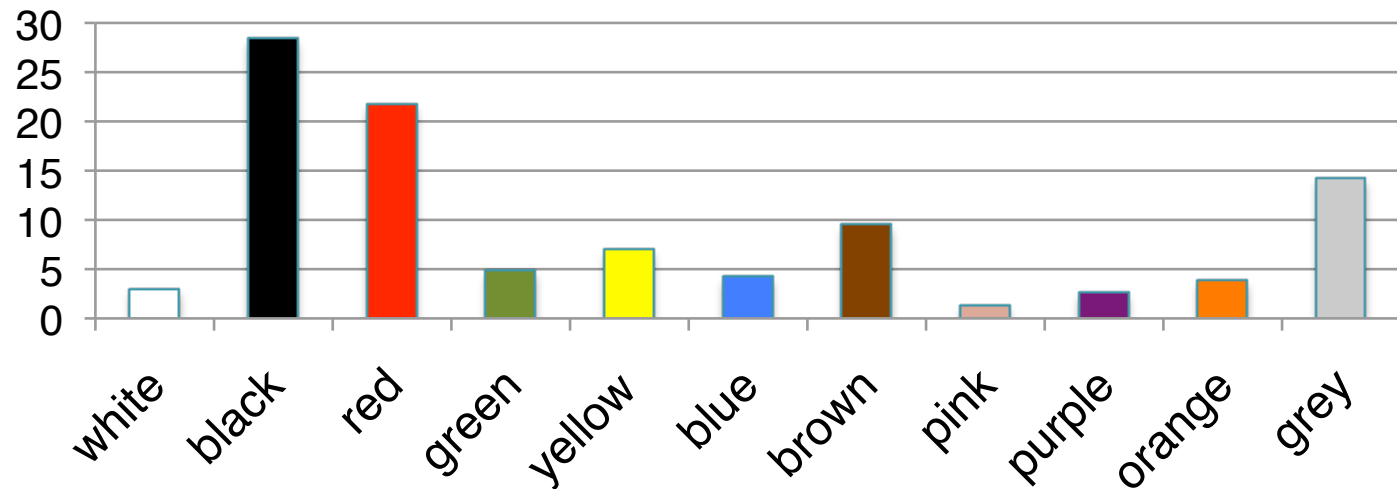
Results

Accuracy, in %

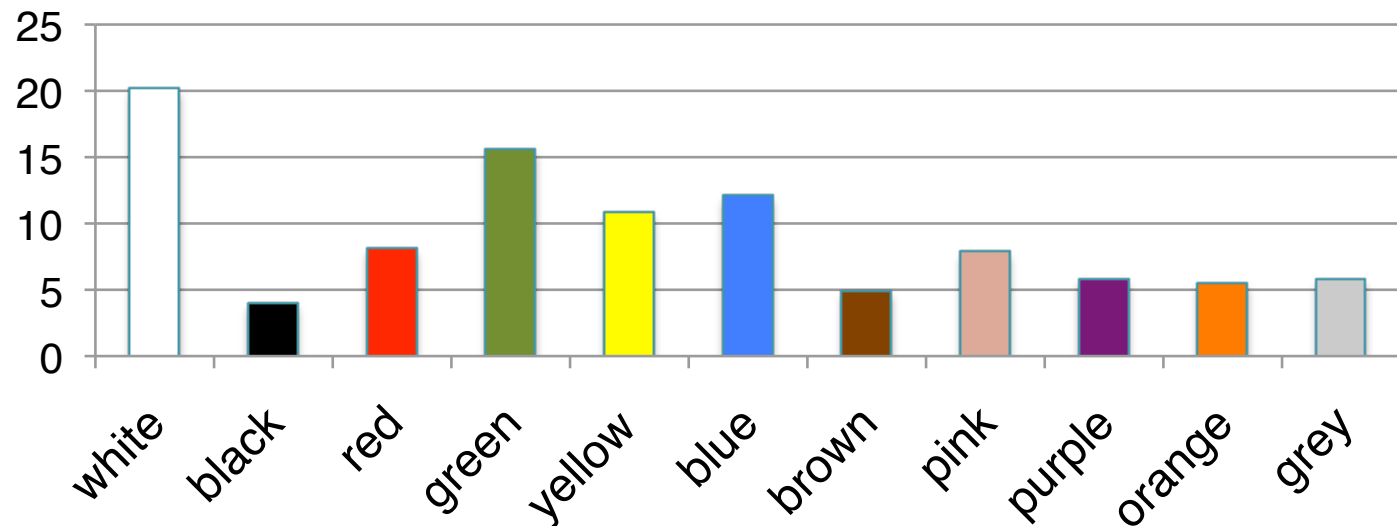


- Above baselines, but not by that much.
- Can polarity help?

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



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

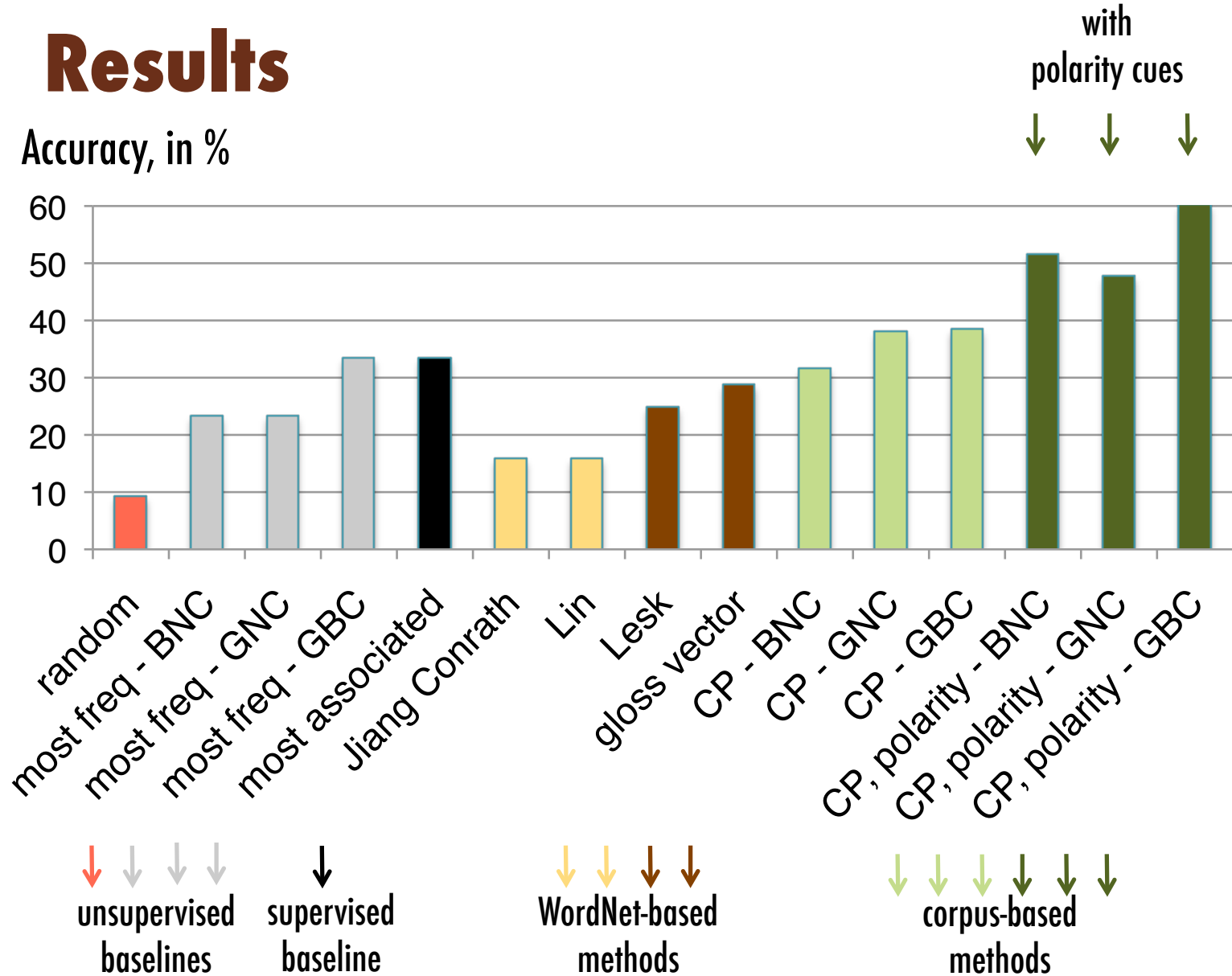


Polarity Cues

- 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 %



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