Words:
Evaluative, Emotional, Colourful, Musical! ♪

Saif Mohammad

National Research Council Canada

Includes joint work with Peter Turney, Tony Yang, Svetlana Kiritchenko, Xiaodan Zhu, Hannah Davis, Colin Cherry, Chris Collins, and Chris Kim.

Word Associations

Beyond literal meaning, words have other associations that often add to their meanings.

- Associations with sentiment
- Associations with emotions
- Associations with social overtones
- Associations with cultural implications
- Associations with colours
- Associations with music

Word Associations

Beyond literal meaning, words have other associations that often add to their meanings.

- Associations with sentiment
- Associations with emotions
- Associations with social overtones
- Associations with cultural implications
- Associations with colours
- Associations with music

Connotations (or all things non-Ptydepe)

Words:

Evaluative, Emotional, Colourful, Musical! ♪

Word-Sentiment Associations

- Adjectives
  - reliable and stunning are typically associated with positive sentiment
  - rude, and broken are typically associated with negative sentiment
- Nouns and verbs
  - surfing and promotion typically associated positive sentiment
  - death and crying typically associated with negative sentiment

Goal: Create a large word-sentiment association lexicon.

Word-Emotion Associations

Words have associations with emotions:

- attack and public speaking typically associated with fear
- yummy and vacation typically associated with joy
- loss and crying typically associated with sadness
- result and wait typically associated anticipation

Goal: Create large word-emotion association lexicon.

Sentiment Analysis

- Is a given sentence positive, negative, or neutral?
- Is a word within a sentence positive, negative, or neutral?

Sentiment Analysis

- Is a given sentence positive, negative, or neutral?
- Is a word within a sentence positive, negative, or neutral?

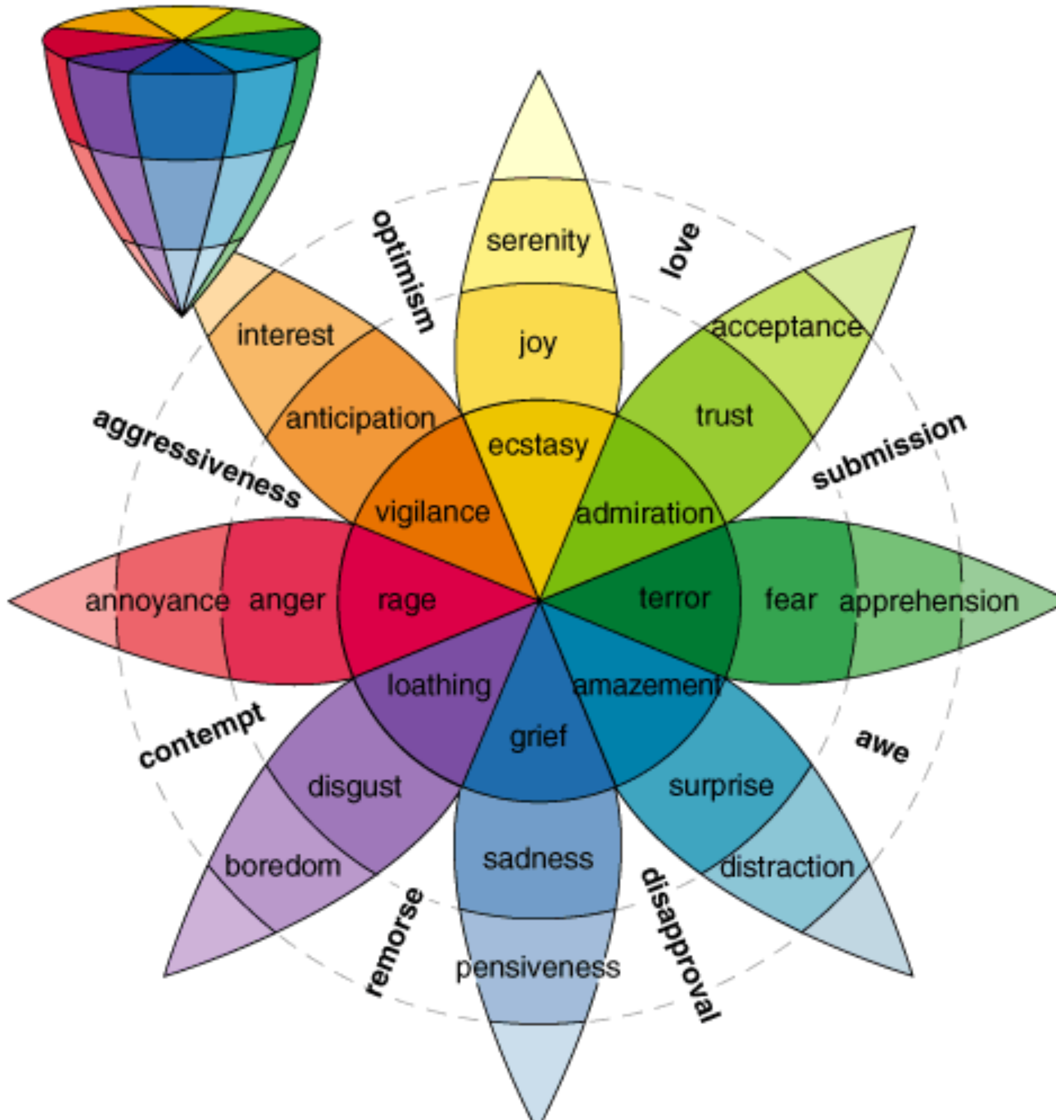
Emotion Analysis

- What emotion is being expressed in a given sentence?
  - joy, sadness, fear, anger, guilt, pride, optimism,…

Which Emotions?

Plutchik, 1980: Eight Basic Emotions

- Joy
- Trust
- Fear
- Surprise
- Sadness
- Disgust
- Anger
- Anticipation



Crowdsourcing

- Benefits
  - Inexpensive
  - Convenient and time-saving
    - Especially for large-scale annotation
- Challenges
  - Quality control
    - Malicious annotations
    - Inadvertent errors
  - Words used in different senses are associated with different emotions.

Word-Choice Question

Peter Turney, NRC

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

- *car*
- *tree*
- *tears*
- *olive*

- Generated automatically
  - Near-synonym taken from thesaurus
  - Distractors are randomly chosen
- Guides Turkers to desired sense
- Aids quality control
  - If Q1 is answered incorrectly:
    - Responses to the remaining questions for the word are discarded

Association Questions

Q2. How much is *cry* associated with the emotion sadness?

(for example, *death* and *gloomy* are strongly associated with sadness)

- *cry* is not associated with sadness
- *cry* is weakly associated with sadness
- *cry* is moderately associated with sadness
- *cry* is strongly associated with sadness

- Eight such questions for the eight basic emotions.
- Two such questions for positive or negative sentiment.

Emotion Lexicon

- Each word-sense pair is annotated by 5 Turkers
- NRC Emotion Lexicon
  - sense-level lexicon
    - word sense pairs: 24,200
  - word-level lexicon
    - union of emotions associated with different senses
    - word types: 14,200

Available at: www.saifmohammad.com

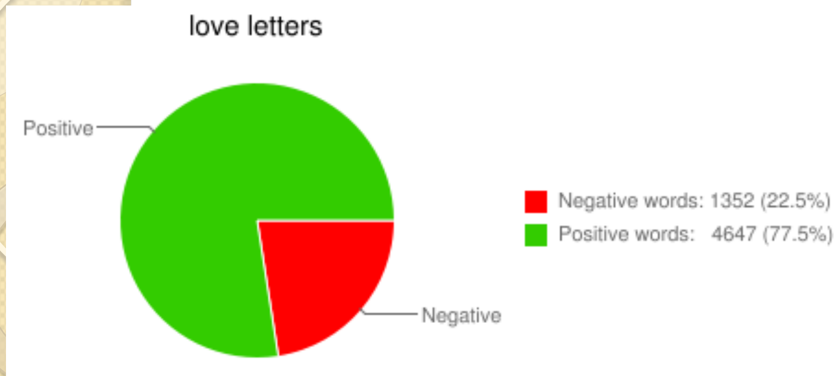
Crowdsourcing a Word-Emotion Association Lexicon, Saif Mohammad and Peter Turney, *Computational Intelligence*, 29 (3), pages 436-465, 2013.

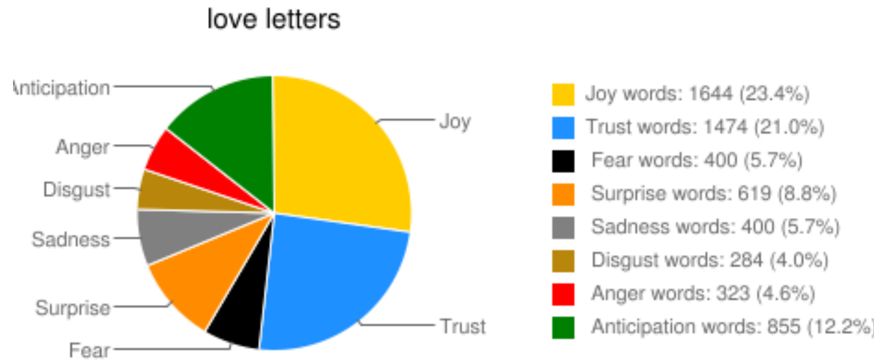
Tony Yang, Simon Fraser University

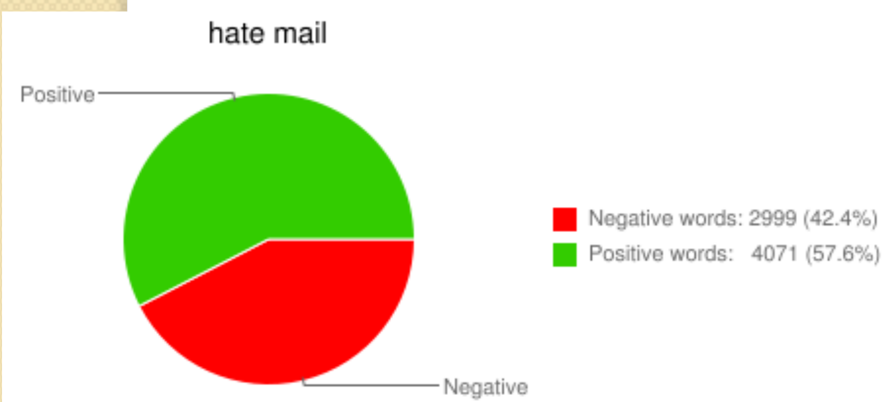
Visualizing Emotions in Text

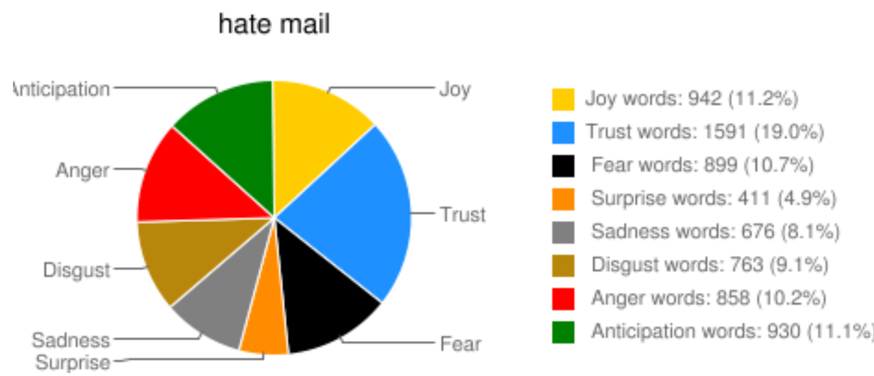
Papers:

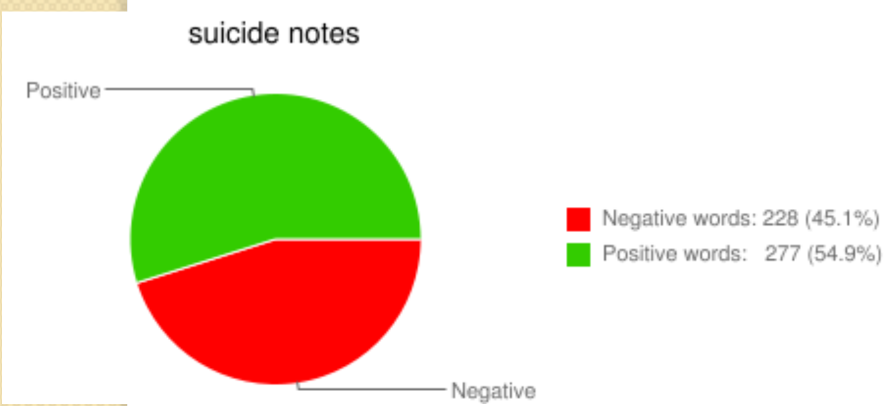
- Tracking Sentiment in Mail: How Genders Differ on Emotional Axes, Saif Mohammad and Tony Yang, In Proceedings of the ACL 2011 Workshop on ACL 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), June 2011, Portland, OR.
- From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales, Saif Mohammad, In Proceedings of the ACL 2011 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH), June 2011, Portland, OR.

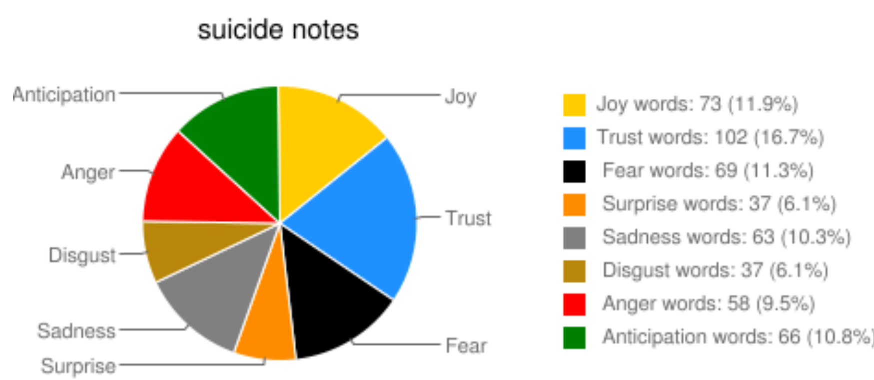

love letters
Positive
Negative
Negative words: 1352 (22.5%)
Positive words: 4647 (77.5%)

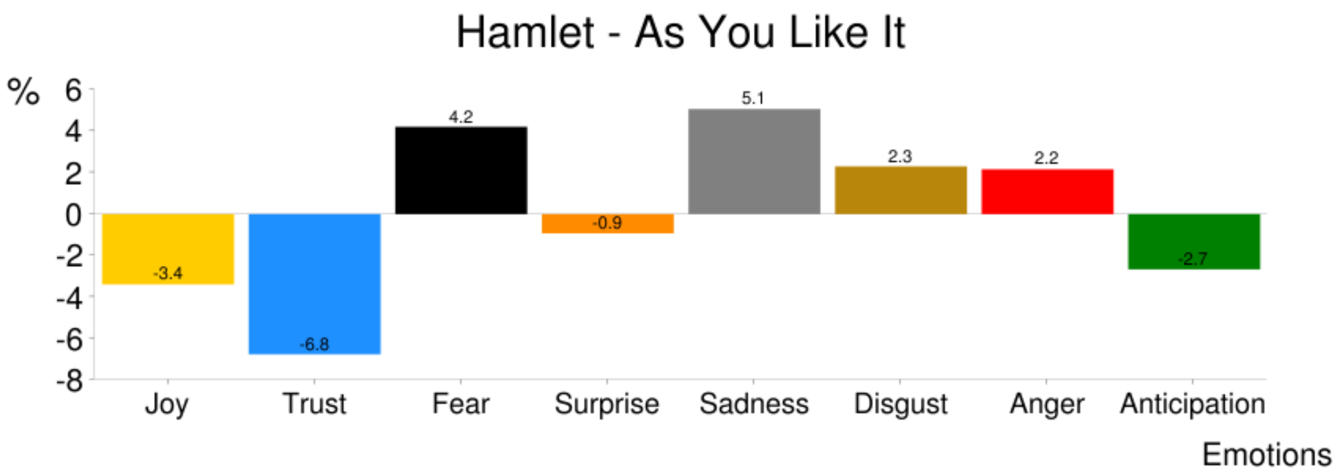

love letters
nticipation
Anger
Disgust
Sadness
Surprise
Fear
Joy
Trust
Joy words: 1644 (23.4%)
Trust words: 1474 (21.0%)
Fear words: 400 (5.7%)
Surprise words: 619 (8.8%)
Sadness words: 400 (5.7%)
Disgust words: 284 (4.0%)
Anger words: 323 (4.6%)
Anticipation words: 855 (12.2%)


hate mail
Positive
Negative
Negative words: 2999 (42.4%)
Positive words: 4071 (57.6%)


hate mail
nticipation
Anger
Disgust
Sadness
Surprise
Joy
Trust
Fear
Joy words: 942 (11.2%)
Trust words: 1591 (19.0%)
Fear words: 899 (10.7%)
Surprise words: 411 (4.9%)
Sadness words: 676 (8.1%)
Disgust words: 763 (9.1%)
Anger words: 858 (10.2%)
Anticipation words: 930 (11.1%)

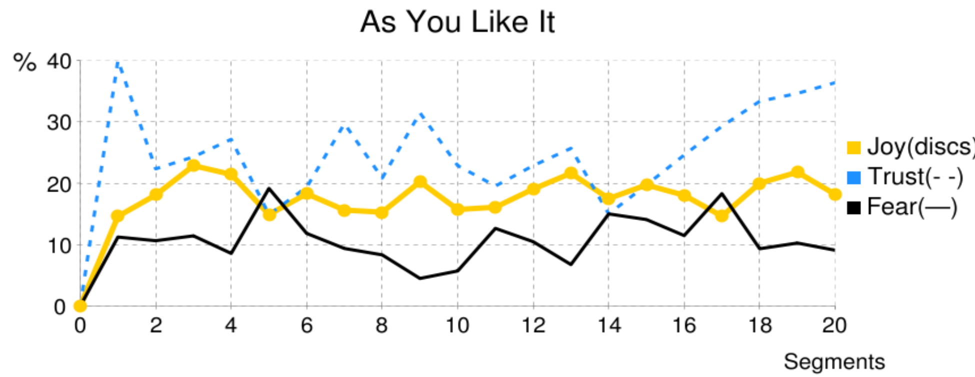

suicide notes
Positive
Negative
Negative words: 228 (45.1%)
Positive words: 277 (54.9%)

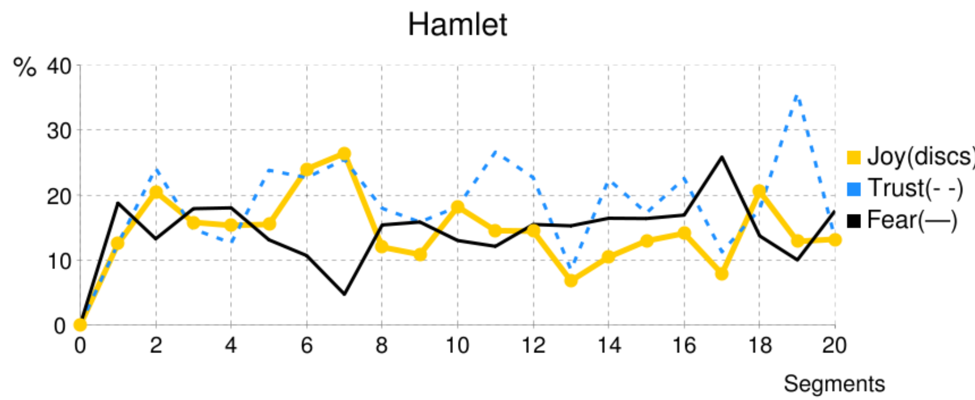

suicide notes
Anticipation
Anger
Disgust
Sadness
Surprise
Joy
Trust
Fear
Joy words: 73 (11.9%)
Trust words: 102 (16.7%)
Fear words: 69 (11.3%)
Surprise words: 37 (6.1%)
Sadness words: 63 (10.3%)
Disgust words: 37 (6.1%)
Anger words: 58 (9.5%)
Anticipation words: 66 (10.8%)

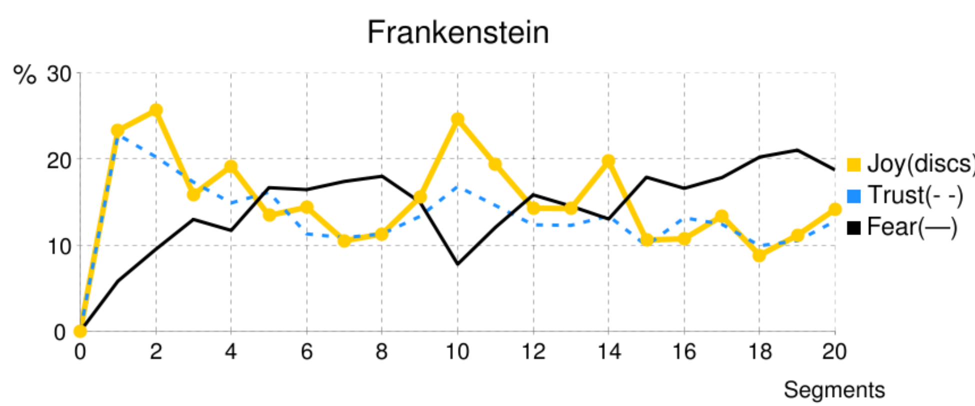


relative salience of trust words

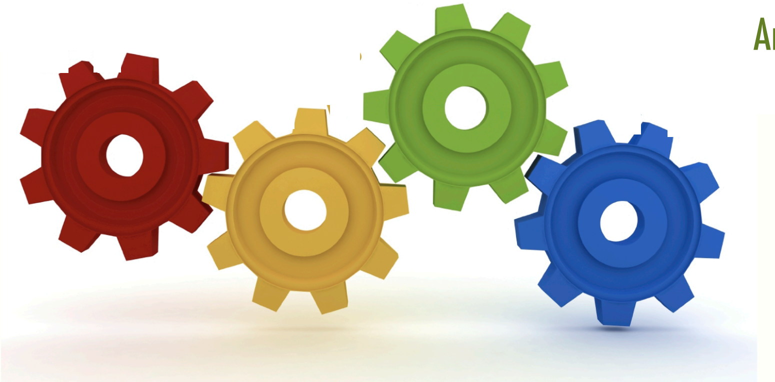
relative salience of sadness words


As You Like It
%
40
30
20
10
0
0 2 4 6 8 10 12 14 16 18 20
Segments
Joy(discs)
Trust(- -)
Fear(—)


Hamlet
%
40
30
20
10
0
0 2 4 6 8 10 12 14 16 18 20
Segments
Joy(discs)
Trust(- -)
Fear(—)

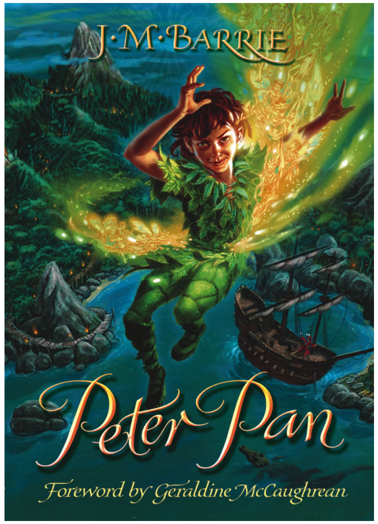

Frankenstein
%
30
20
10
0
0 2 4 6 8 10 12 14 16 18 20
Segments
Joy(discs)
Trust(- -)
Fear(—)

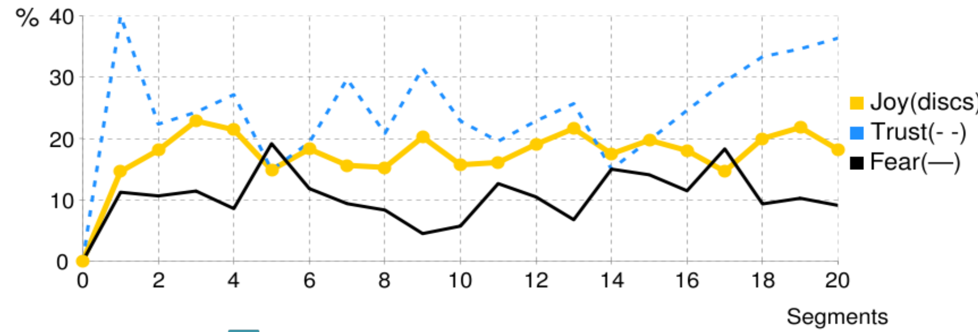
Hannah Davis
Artist/Programmer

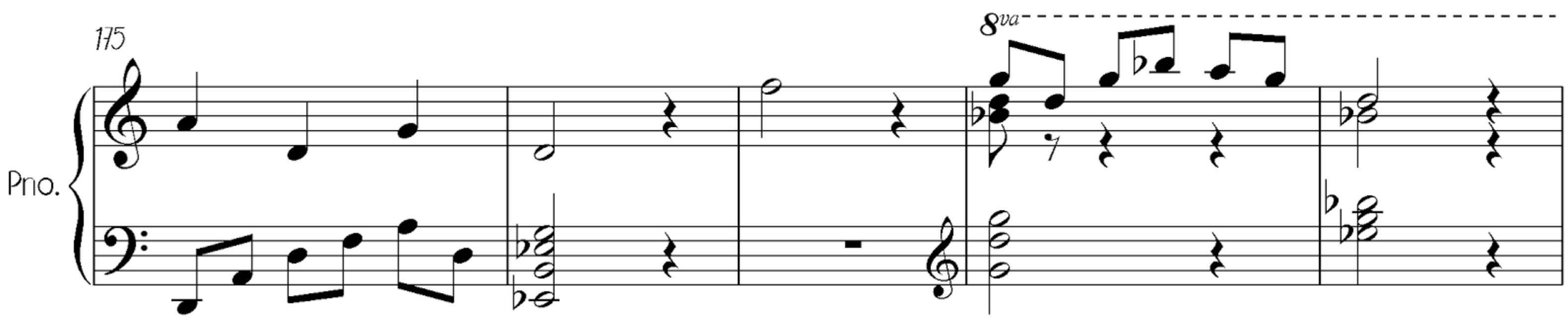


Words:
Evaluative, Emotional, Colourful, Musical! ♪

- Generating Music from Literature. Hannah Davis and Saif M. Mohammad, In Proceedings of the EACL Workshop on Computational Linguistics for Literature, April 2014, Gothenburg, Sweden.







A method to generate music from literature.

- music that captures the change in the distribution of emotion words.

Challenges

- Not changing existing music -- generating novel pieces
- Paralysis of choice
- Has to sound good
- No one way is the right way
- Evaluation is tricky

Music-Emotion Associations

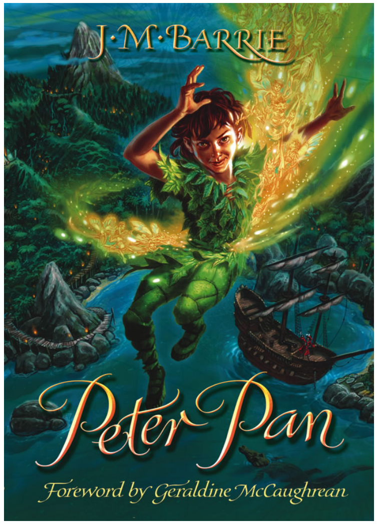
- Major and Minor Keys
  - major keys: happiness
  - minor keys: sadness
- Tempo
  - fast tempo: happiness or excitement
- Melody
  - a sequence of consonant notes: joy and calm
  - a sequence of dissonant notes: excitement, anger, or unpleasantness

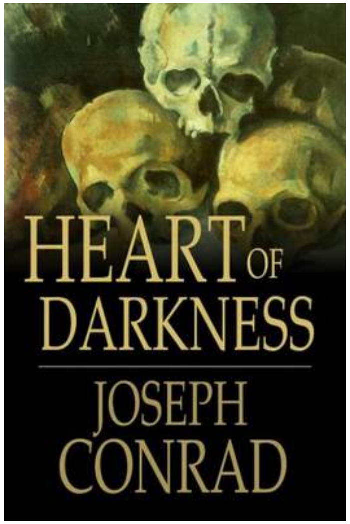
Hunter et al., 2010, Hunter et al., 2008, Ali and Peynirciolu, 2010, Gabrielsson and Lindstrom, 2001, Webster and Weir, 2005

Pieces

Three simultaneous piano melodies pertaining to the dominant emotions.

Examples (click cover to play)





TransProse: www.musicfromtext.com
Music played 185,000 times in the last 2 months.

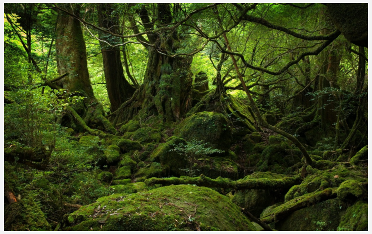
Words:

Evaluative, Emotional, Colourful, Musical!♪

Word-Colour Associations

Concrete concepts

iceberg → white



vegetation → green

Abstract concepts

danger → red

honesty → white

Colours Add to Linguistic Information

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

Crowdsourced Questionnaire

Q. Which colour is associated with *sleep*?

- black • green • purple… (11 colour options in random order)

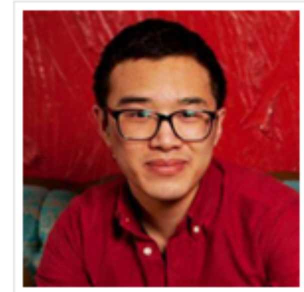
NRC Word-Colour Association Lexicon

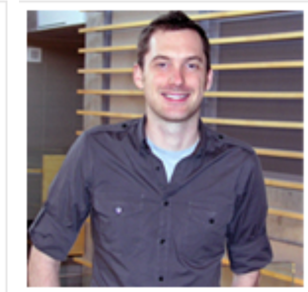
- sense-level lexicon: 24,200 word sense pairs

Papers:

- Colourful Language: Measuring Word-Colour Associations, Saif Mohammad, In Proceedings of the ACL 2011 Workshop on Cognitive Modeling and Computational Linguistics (CMCL), June 2011, Portland, OR.
- Even the Abstract have Colour: Consensus in Word-Colour Associations, Saif Mohammad, In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, June 2011, Portland, OR.

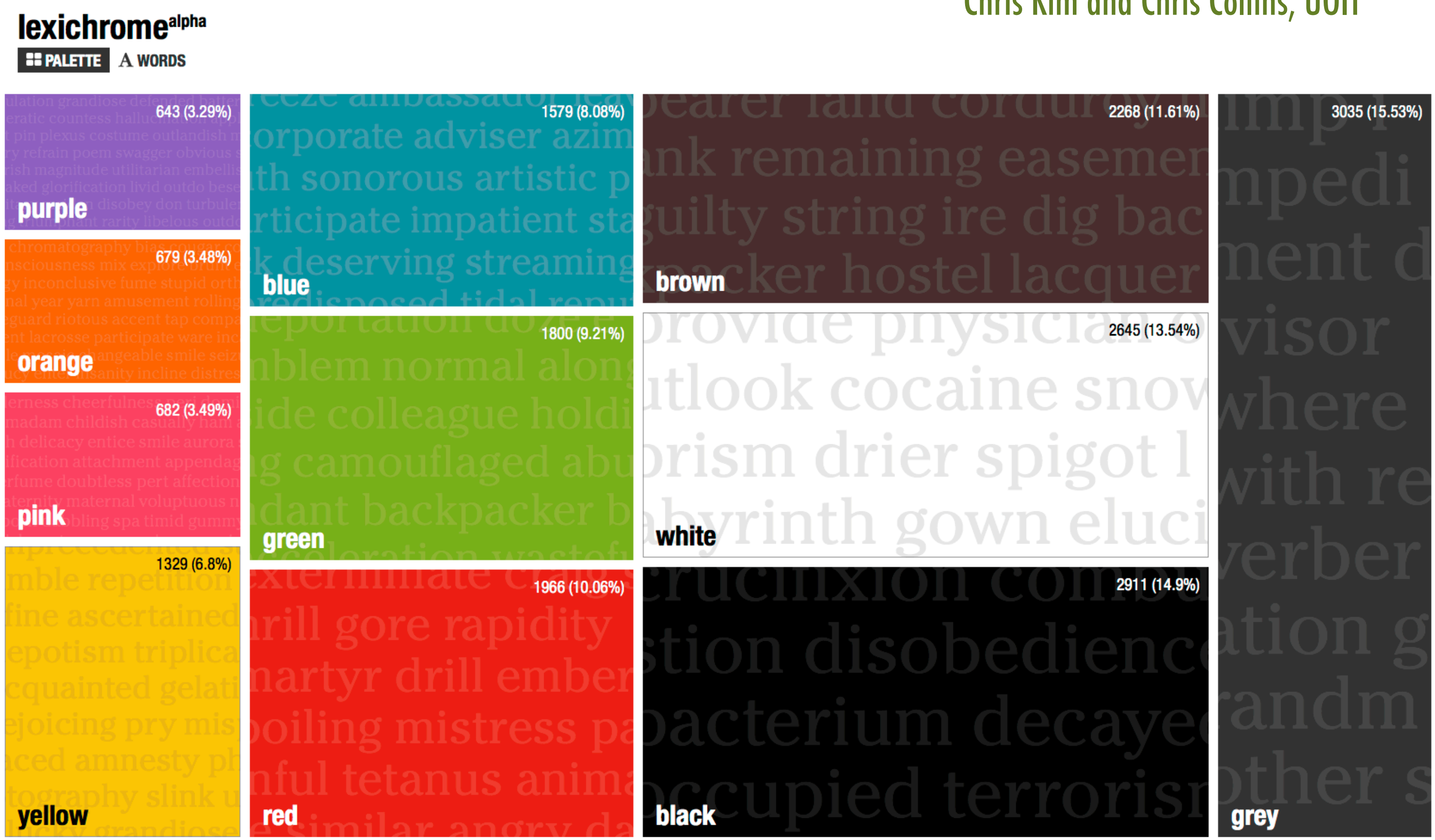
Visualizing Word-Colour Associations





Chris Kim and Chris Collins, UOIT

[Visualization]



‹ all words associated with green

PALETTE | WORDS

RELEVANCE (DESC) | ALPHABETICAL

botany 8 out of 8	evergreen 12 out of 12	garden 12 out of 12	pickle 8 out of 8	sprout 13 out of 13	vegetable 8 out of 8
moss 14 out of 15	weed 9 out of 10	herbal 8 out of 9	remittance 8 out of 9	mint 15 out of 17	frog 14 out of 16
habitat 7 out of 8	meadow 7 out of 8	monetary 7 out of 8	pay 7 out of 8	wealth 9 out of 11	plantation 8 out of 10
abundance 6 out of 8	cost 12 out of 16	dollar 6 out of 8	endeavor 6 out of 8	grantee 6 out of 8	leaflet 6 out of 8
propagation 9 out of 12	worth 9 out of 12	mow 17 out of 23	leaf 14 out of 19	farm 16 out of 22	garnish 16 out of 22
amount 22 out of 31	greedy 7 out of 10	cultivation 6 out of 9	financier 6 out of 9	germ 10 out of 15	lavish 6 out of 9
payment 8 out of 12	pod 6 out of 9	price 14 out of 21	swampy 6 out of 9	weeds 17 out of 26	affluence 15 out of 23
harvest 9 out of 14	restitution 7 out of 11	alienation 10 out of 16	army 25 out of 40	camouflage 5 out of 8	gain 5 out of 8
gainful 5 out of 8	graze 5 out of 8	renewal 5 out of 8	teller 5 out of 8	vineyard 5 out of 8	wreath 5 out of 8
buyer	debenture	inhabit	assets	appraise	envious

< all words associated with black

PALETTE A WORDS

RELEVANCE (DESC) ALPHABETICAL

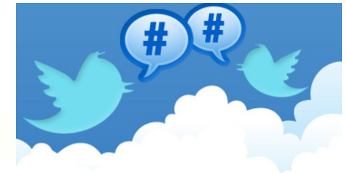
blackness 14 out of 15	black 13 out of 14	evil 12 out of 13	thug 13 out of 15	darken 28 out of 33	charcoal 25 out of 30
mourn 10 out of 12	interment 9 out of 11	negro 8 out of 10	suicidal 8 out of 10	pepper 10 out of 13	death 9 out of 12
somber 14 out of 19	bomb 16 out of 22	executioner 8 out of 11	perish 16 out of 22	subversion 8 out of 11	blindfold 10 out of 14
curse 17 out of 24	sin 19 out of 27	recording 7 out of 10	stormy 7 out of 10	vulture 7 out of 10	cannon 9 out of 13
nocturnal 9 out of 13	ominous 9 out of 13	soot 9 out of 13	witch 9 out of 13	adversity 11 out of 16	discrimination 15 out of 22
marked 8 out of 12	sinister 8 out of 12	downfall 21 out of 32	grieve 21 out of 32	dire 11 out of 17	dislike 11 out of 17
illegal 9 out of 14	schism 9 out of 14	fright 7 out of 11	terminal 7 out of 11	threatening 7 out of 11	despair 27 out of 43
decayed 10 out of 16	print 15 out of 24	abomination 8 out of 13	abyss 16 out of 26	aversion 8 out of 13	contraband 8 out of 13
deadly 11 out of 18	disastrous 14 out of 23	decimal 6 out of 10	deepen 6 out of 10	demise 9 out of 15	desolation 6 out of 10
hag 6 out of 10	missing 9 out of 15	prostitute 6 out of 10	reprint 6 out of 10	suffering 6 out of 10	thief 12 out of 20

< all words associated with yellow

PALETTE A WORDS

RELEVANCE (DESC) ALPHABETICAL

cowardly 10 out of 10	nugget 7 out of 7	sun 7 out of 7	sunny 9 out of 10	saffron 8 out of 9	treasure 7 out of 8
lion 6 out of 7	mustard 6 out of 7	radiant 6 out of 7	bee 11 out of 13	butter 11 out of 13	insecure 6 out of 8
sandy 6 out of 8	scatter 6 out of 8	lightning 8 out of 11	beehive 10 out of 14	practically 5 out of 7	radiate 5 out of 7
enlighten 7 out of 10	sunshine 7 out of 10	candlelight 6 out of 9	dawn 4 out of 6	honey 6 out of 9	urinalysis 6 out of 9
honeycomb 9 out of 14	awaken 7 out of 11	cornet 5 out of 8	daze 5 out of 8	kerosene 5 out of 8	oriental 5 out of 8
phosphor 5 out of 8	pyramid 10 out of 16	quail 5 out of 8	sand 5 out of 8	omelet 6 out of 10	breakfast 7 out of 12
medal 14 out of 24	acquainted 4 out of 7	candle 4 out of 7	chirp 8 out of 14	conversion 4 out of 7	egg 4 out of 7
incandescent 4 out of 7	innuendo 4 out of 7	lessen 4 out of 7	bugle 5 out of 9	bus 5 out of 9	coin 5 out of 9
day 5 out of 9	folly 5 out of 9	hive 5 out of 9	insignificant 5 out of 9	lighthouse 10 out of 18	plated 5 out of 9
ripple 5 out of 9	beam 6 out of 11	happy 7 out of 13	buzz 6 out of 12	caution 4 out of 8	conductivity 4 out of 8



Words:

Evaluative, Emotional, Colourful, Musical!

Hashtagged Tweets

- Hashtagged words are good labels of sentiments and emotions

  Some jerk just stole my photo on #tumblr #grrr **#anger**

- Hashtags are not always good labels:
  - hashtag used sarcastically

    The reviewers want me to re-annotate the data. **#joy**
  - hashtagged emotion not in the rest of the message

    Mika used my photo on tumblr. **#anger**

#Emotional Tweets, Saif Mohammad, In Proceedings of the First Joint Conference on Lexical and Computational Semantics (*Sem), June 2012, Montreal, Canada.

Generating lexicon for 500 emotions

NRC Hashtag Emotion Lexicon: About 20,000 words associated with about 500 emotions

Papers:

- Using Nuances of Emotion to Identify Personality. Saif M. Mohammad and Svetlana Kiritchenko, In Proceedings of the ICWSM Workshop on Computational Personality Recognition, July 2013, Boston, USA.
- Using Hashtags to Capture Fine Emotion Categories from Tweets. Saif M. Mohammad, Svetlana Kiritchenko, Computational Intelligence, in press.

SemEval-2013, Task 2

Svetlana Kiritchenko
NRC

Xiaodan Zhu
NRC

- Is a given message positive, negative, or neutral?
  - tweet or SMS
- Is a given term within a message positive, negative, or neutral?

International competition on sentiment analysis of tweets:

- SemEval-2013 (co-located with NAACL-2013)
- 44 teams

NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets, Saif M. Mohammad, Svetlana Kiritchenko, and Xiaodan Zhu, In Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013), June 2013, Atlanta, USA.

Sentiment Lexicons

Lists of positive and negative words.

Positive
spectacular
okay

Negative
lousy
unpredictable

Sentiment Lexicons

Lists of positive and negative words, with scores indicating the degree of association

Positive
spectacular 0.91
okay 0.3

Negative
lousy -0.84
unpredictable -0.17

Creating New Sentiment Lexicons

- Compiled a list of seed sentiment words by looking up synonyms of excellent, good, bad, and terrible:
  - 30 positive words
  - 46 negative words
- Polled the Twitter API for tweets with seed-word hashtags
  - A set of 775,000 tweets was compiled from April to December 2012

Automatically Generated New Lexicons

- A tweet is considered:
  - positive if it has a positive hashtag
  - negative if it has a negative hashtag

- For every word w in the set of 775,000 tweets, an association score is generated:

$$score(w) = PMI(w, positive) - PMI(w, negative)$$

PMI = pointwise mutual information

If $score(w) > 0$, then w is positive

If $score(w) < 0$, then w word is negative

NRC Hashtag Sentiment Lexicon

- w can be:
  - any unigram in the tweets: 54,129 entries
  - any bigram in the tweets: 316,531 entries
  - non-contiguous pairs (any two words) from the same tweet: 308,808 entries

- Multi-word entries incorporate context:

  unpredictable story 0.4

  unpredictable steering -0.7

Setup

- **Pre-processing:**
  - URL -> http://someurl
  - UserID -> @someuser
  - Tokenization and part-of-speech (POS) tagging (CMU Twitter NLP tool)
- **Classifier:**
  - SVM with linear kernel
- **Evaluation:**
  - Macro-averaged F-pos and F-neg

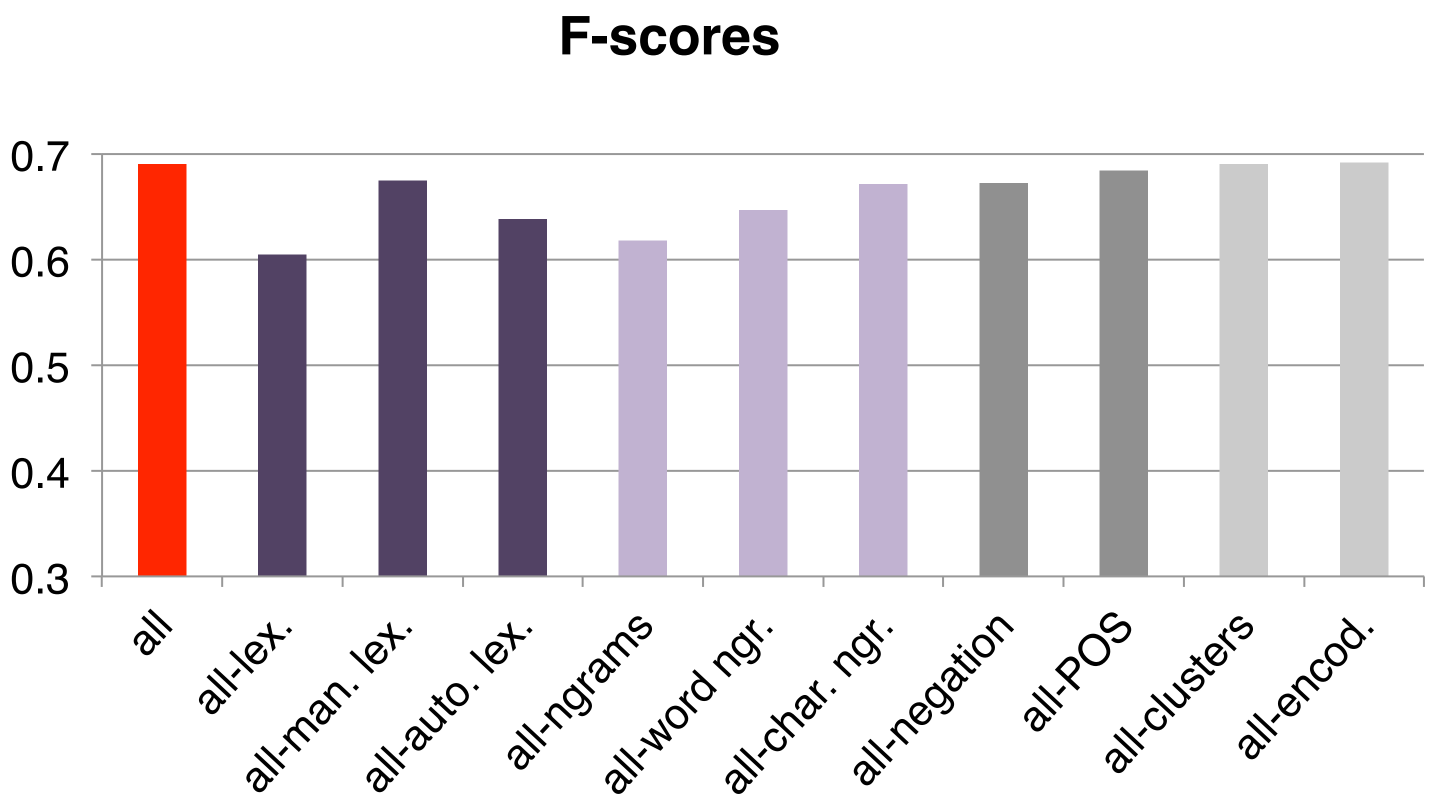
Features

Features	Examples
sentiment lexicon	#positive: 3, scorePositive: 2.2; maxPositive: 1.3; last: 0.6, scoreNegative: 0.8, scorePositive_neg: 0.4
word n-grams	spectacular, like documentary
char n-grams	spect, docu, visua
part of speech	#N: 5, #V: 2, #A:1
negation	#Neg: 1; ngram:perfect → ngram:perfect_neg, polarity:positive → polarity:positive_neg
word clusters	probably, definitely, def
all-caps	YES, COOL
punctuation	#!+: 1, #?+: 0, #!?+: 0
word clusters	probably, definitely, probly
emoticons	:D, >:(
elongated words	soooo, yaayyy

NRC-Canada's Rankings in SemEval Tasks

- SemEval-2013 Task 2: Sentiment Analysis in Twitter (40+ teams)
  - Tweets
    - message-level: 1st rank
    - term level: 1st rank
  - SMS messages
    - message-level: 1st rank
    - term level: 2nd rank
- SemEval-2014 Task 9: Sentiment Analysis in Twitter (40+ teams)
  - 1st rank in 5 of 10 subtask-dataset combinations
- SemEval-2014 Task 4: Aspect Based Sentiment Analysis (30+ teams)
  - 1st rank in two of the three sentiment subtasks and 2nd in the other

Feature Contributions (on Tweets)



Movie Reviews

- Data from rottentomatoes.com (Pang and Lee, 2005)
- Socher et al. (2013) training and test set up
- Message-level task
  - Two-way classification: positive or negative

	System	**Accuracy**
(a)	Majority baseline	50.1
(b)	SVM-unigrams	71.9
(c)	Previous best result (Socher et al., 2013)	85.4
(d)	Our system	**85.5**

Mechanical Turk Annotations

- NRC word-emotion association lexicon
  - Entries for 8 emotion and 2 sentiments
  - Entries for 14,200 words
- NRC word-colour association lexicon
  - Entries for 11 colours
  - Entries for 14,200 words

Automatically Generated Lexicons (from Tweets)

- Hashtag Emotion Lexicon
  - 500 emotions
  - Entries for 20,000 words
- Hashtag Sentiment Lexicon
  - Entries for positive and negative sentiment
  - Entries for 54,000 words

Saif Mohammad

www.saifmohammad.com

saif.mohammad@nrc-cnrc.gc.ca

Future Directions

- Bridging sentiment and emotion analysis with other modalities like music
  - Further developing the text-to-music system
  - Identifying interesting applications
- Using sentiment analysis for related problems such as stance detection
- Working not just with basic emotions but with any of the hundreds of emotions
  - Understanding the relationships between different emotions
- Developing visualizations that help understand word associations