The Words are Alive: Associations with Sentiment, Em☺tion, Colour, and Music ♪

Saif Mohammad

National Research Council Canada

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

National Research Council Canada

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.

The Words are Alive: Associations with Sentiment, Em☺tion, Colour, and Music

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
  - holiday and smiling are typically associated positive sentiment
  - death and crying are typically associated with negative sentiment

Goal: Capture word-sentiment associations.

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: Capture word-emotion associations.

Sentiment Analysis

- Is a given piece of text positive, negative, or neutral?
    - The text may be a sentence, a tweet, an SMS message, a customer review, a document, and so on.

Sentiment Analysis

- Is a given piece of text positive, negative, or neutral?
  - The text may be a sentence, a tweet, an SMS message, a customer review, a document, and so on.

Emotion Analysis

- What emotion is being expressed in a given piece of text?
  - Example emotions: joy, sadness, fear, anger, guilt, pride, optimism, frustration,…

Applications of Sentiment Analysis and Emotion Detection

- Tracking sentiment towards politicians, movies, products
- Improving customer relation models
- Identifying what evokes strong emotions in people
- Detecting happiness and well-being
- Measuring the impact of activist movements through text generated in social media.
- Improving automatic dialogue systems
- Improving automatic tutoring systems
- Detecting how people use emotion-bearing-words and metaphors to persuade and coerce others

Sentiment Analysis: Tasks

- Is a given piece of text positive, negative, or neutral?
  - The text may be a sentence, a tweet, an SMS message, a customer review, a document, and so on.

Sentiment Analysis: Tasks

- Is a given piece of text positive, negative, or neutral?
  - The text may be a sentence, a tweet, an SMS message, a customer review, a document, and so on.
- Is a word within a sentence positive, negative, or neutral?
  - unpredictable movie plot vs. unpredictable steering
- What is the sentiment towards specific aspects of a product?
  - sentiment towards the food and sentiment towards the service in a customer review of a restaurant
- What is the sentiment towards an entity such as a politician, government policy, company, issue, or product.
  - Stance detection: favorable or unfavorable
  - Framing: focusing on specific dimensions

Sentiment Analysis: Tasks (continued)

- What is the sentiment of the speaker/writer?
  - Is the speaker explicitly expressing sentiment?
- What sentiment is evoked in the listener/reader?
- What is the sentiment of an entity mentioned in the text?

Consider the examples below:

Team Tapioca destroyed the BeetRoots.

General Tapioca was killed in an explosion.

General Tapioca was ruthlessly executed today.

Mass-murdered General Tapioca finally found and killed in battle.

May God be with the families of the fallen.

Sentiment Analysis

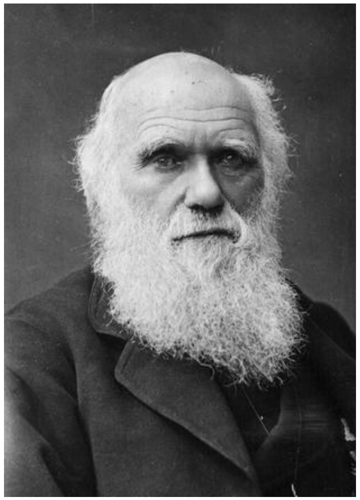
- Is a given piece of text positive, negative, or neutral?
  - The text may be a sentence, a tweet, SMS message, a customer review, a document, and so on.

Emotion Analysis

- What emotion is being expressed in a given piece of text?
  - Example emotions: joy, sadness, fear, anger, guilt, pride, optimism, frustration,…

Which Emotions?

Charles Darwin



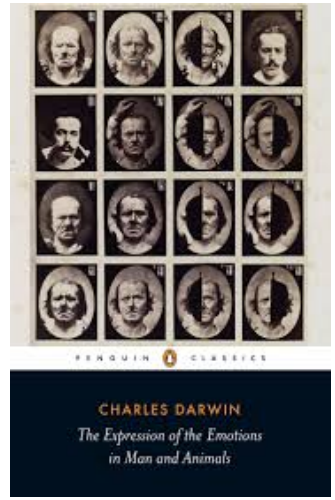
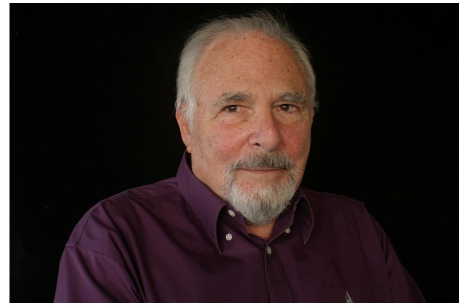


FIG. 20.—Terror,
from a photograph by Dr. Duchenne.

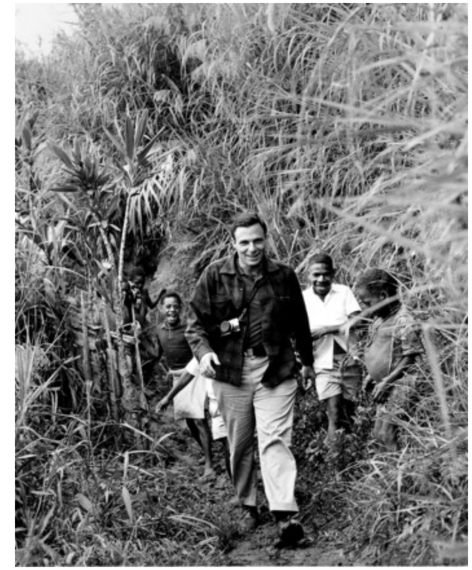
- published *The Expression of the Emotions in Man and Animals* in 1872
- seeks to trace the animal origins of human characteristics
  - pursing of the lips in concentration
  - tightening of the muscles around the eyes in anger
- claimed that certain facial expressions are universal
  - these facial expressions are associated with emotions

Debate: Universality of Perception of Emotions

Margaret Mead
Cultural anthropologist



Paul Ekman
Psychologist and discoverer of micro expressions.



- Circa 1950's, Margaret Mead and others believed facial expressions and their meanings were culturally determined
    - behavioural learning processes
- Paul Ekman provided the strongest evidence to date that Darwin, not Margaret Mead, was correct in claiming facial expressions are universal
- Found universality of six emotions

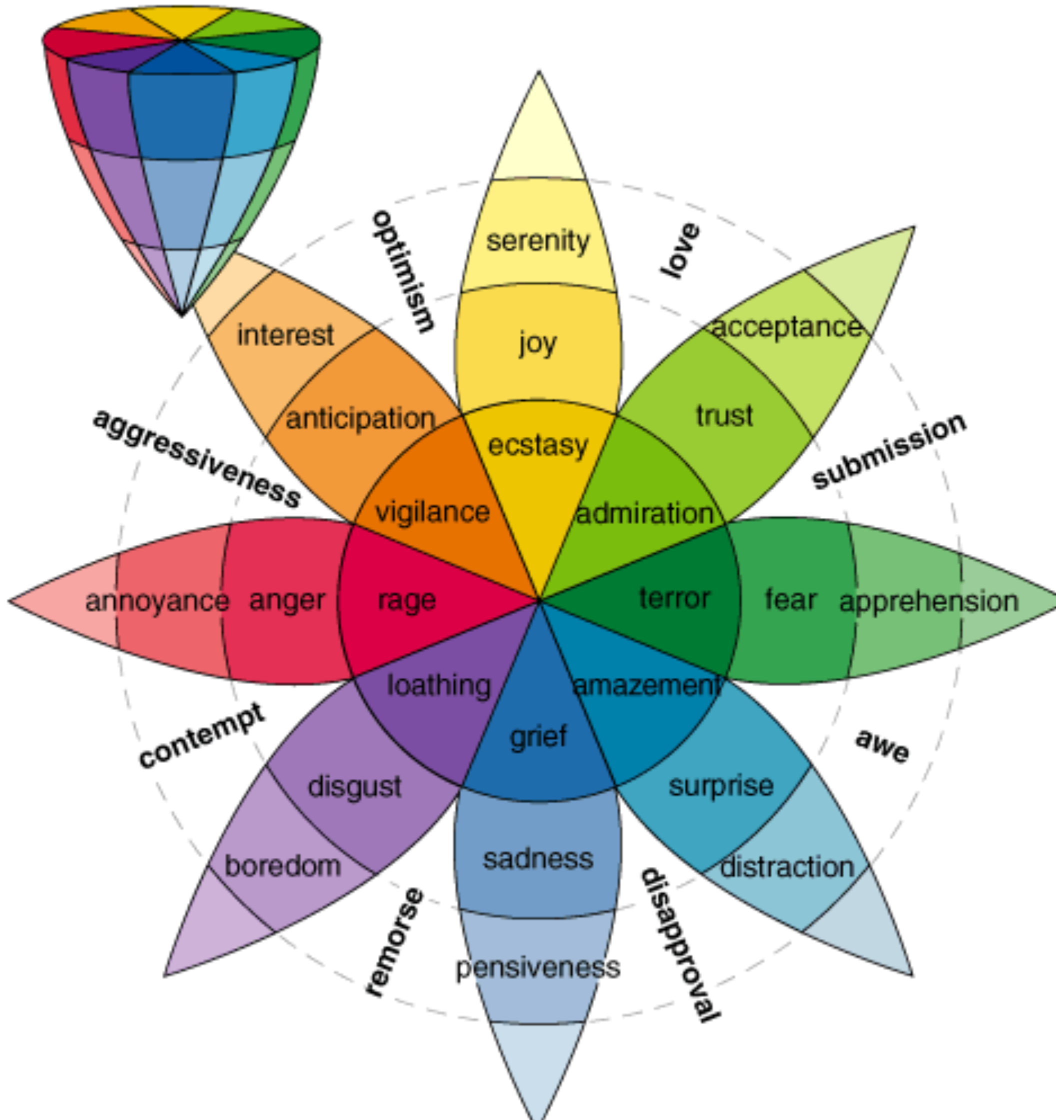
Paul Ekman, 1971: Six Basic Emotions

- Anger
- Disgust
- Fear
- Joy
- Sadness
- Surprise

A collection of six photographs from Paul Ekman's book 'Emotions Revealed.'

Plutchik, 1980: Eight Basic Emotions

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



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: Capture word-emotion associations.

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.

Better agreement when asked ‘associated with’ rather than ‘evoke’.

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

Paper:

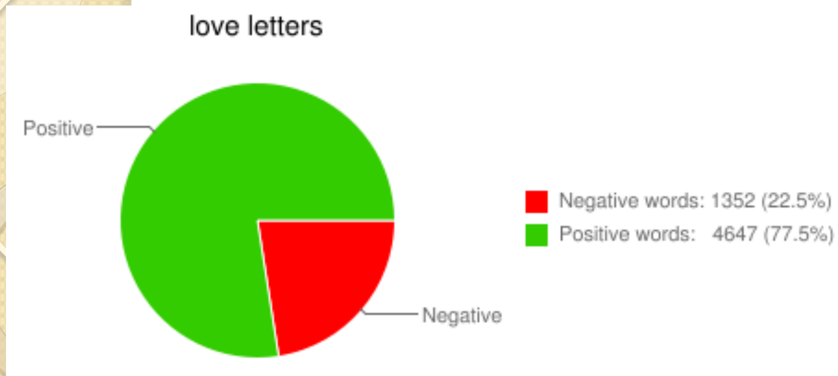
Crowdsourcing a Word-Emotion Association Lexicon, Saif Mohammad and Peter Turney, *Computational Intelligence*, 29 (3), pages 436-465, 2013.

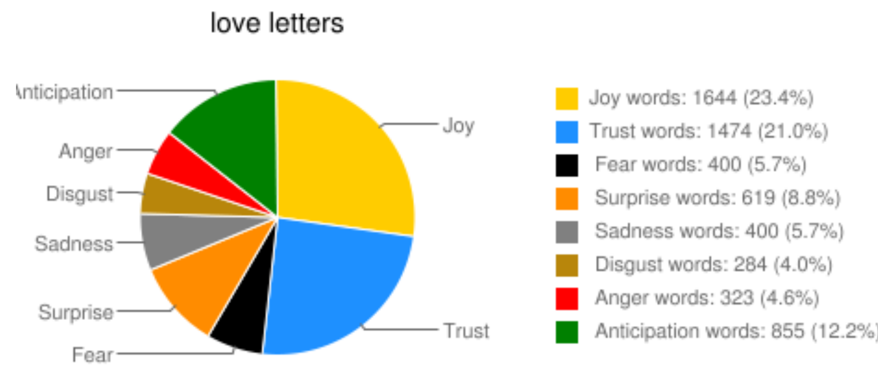
Tony Yang, Simon Fraser University

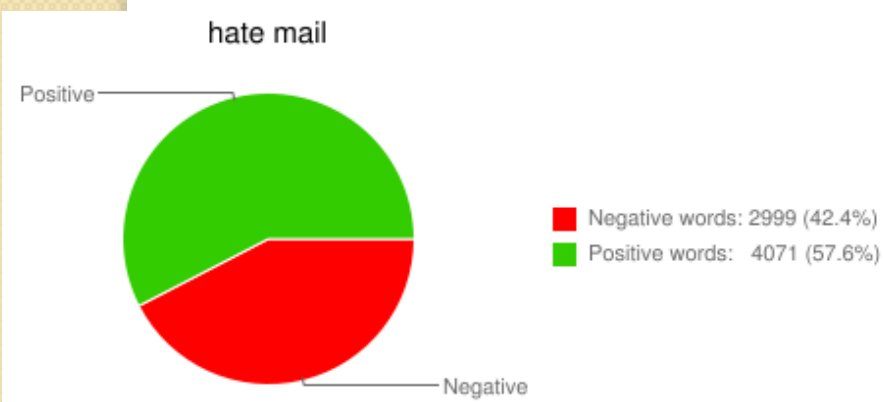
Visualizing Em☺tions in Text

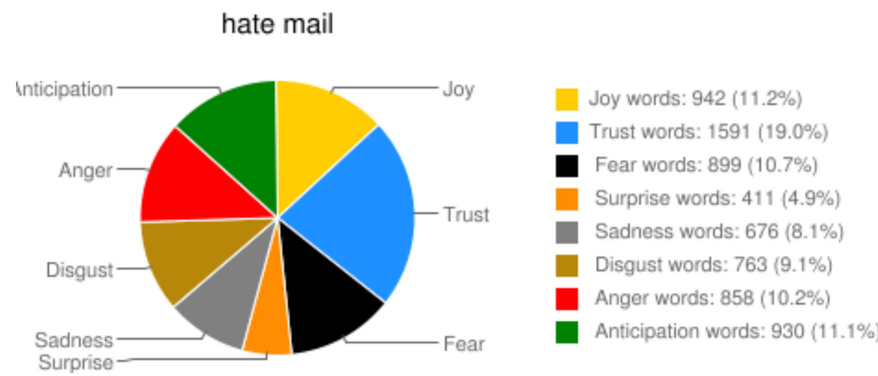
Papers:

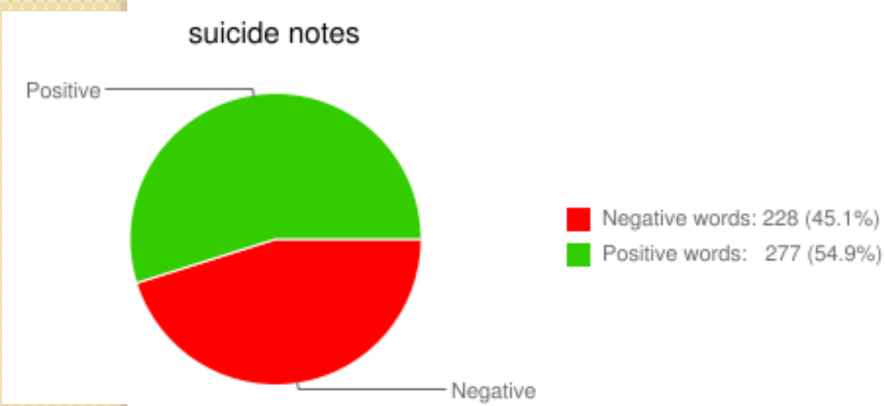
- Tracking Sentiment in Mail: How Genders Differ on Emotional Axes, Saif Mohammad and Tony Yang, In Proceedings of the ACL 2011 Workshop on ACL 2011 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), June 2011, Portland, OR.
- From Once Upon a Time to Happily Ever After: 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.

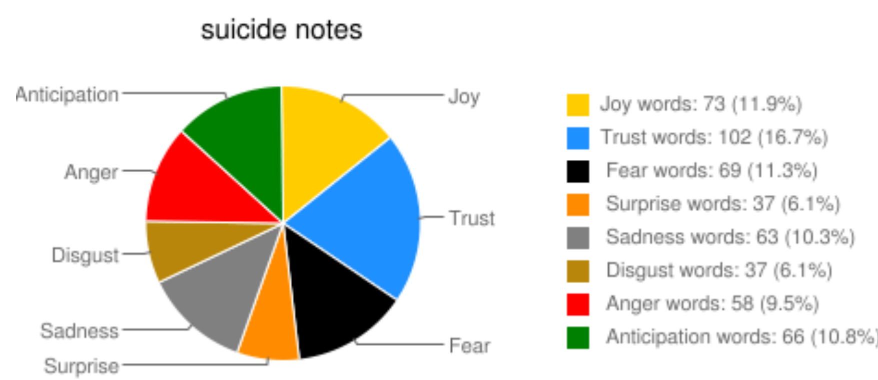




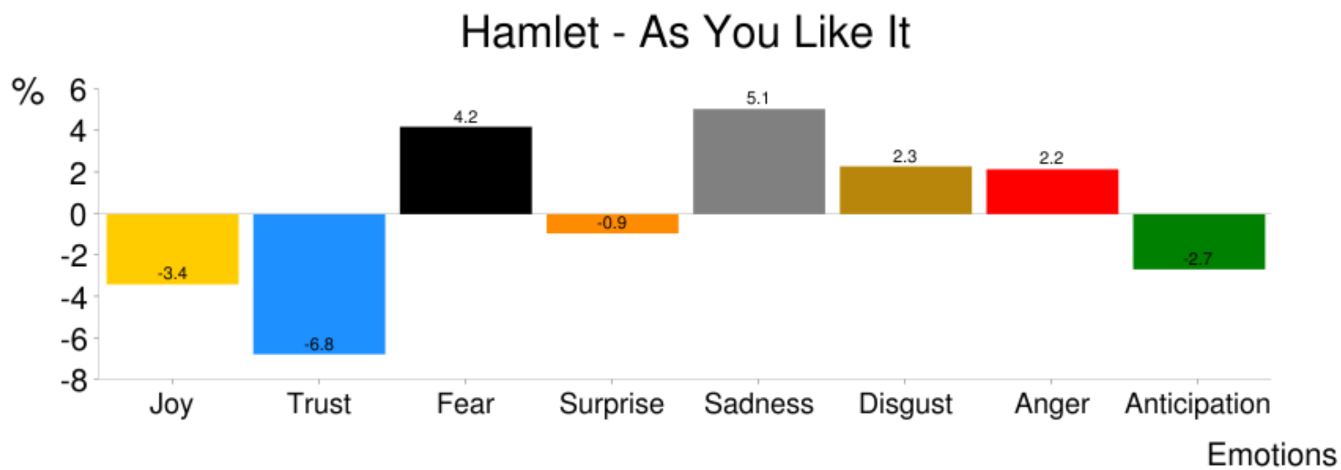








National Research Council Canada

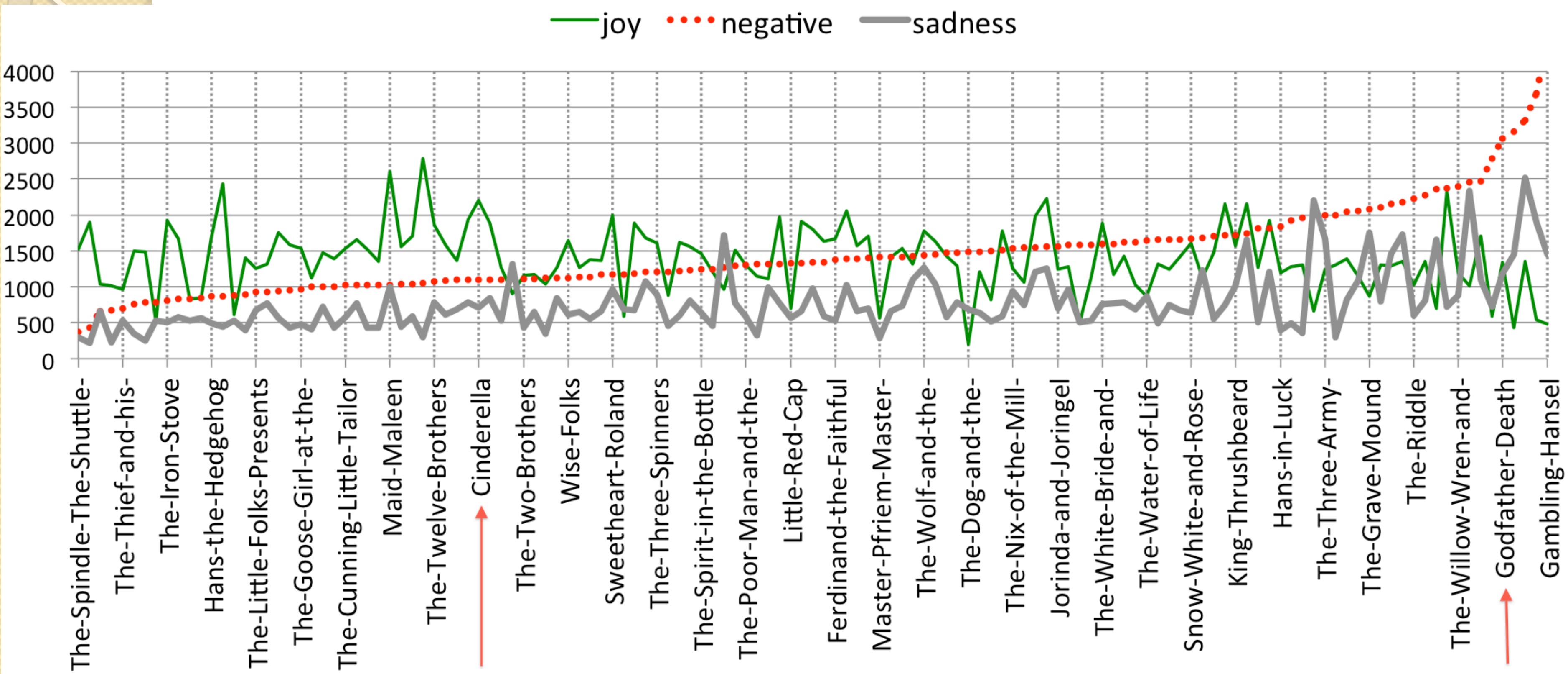


relative salience of trust words

relative salience of sadness words

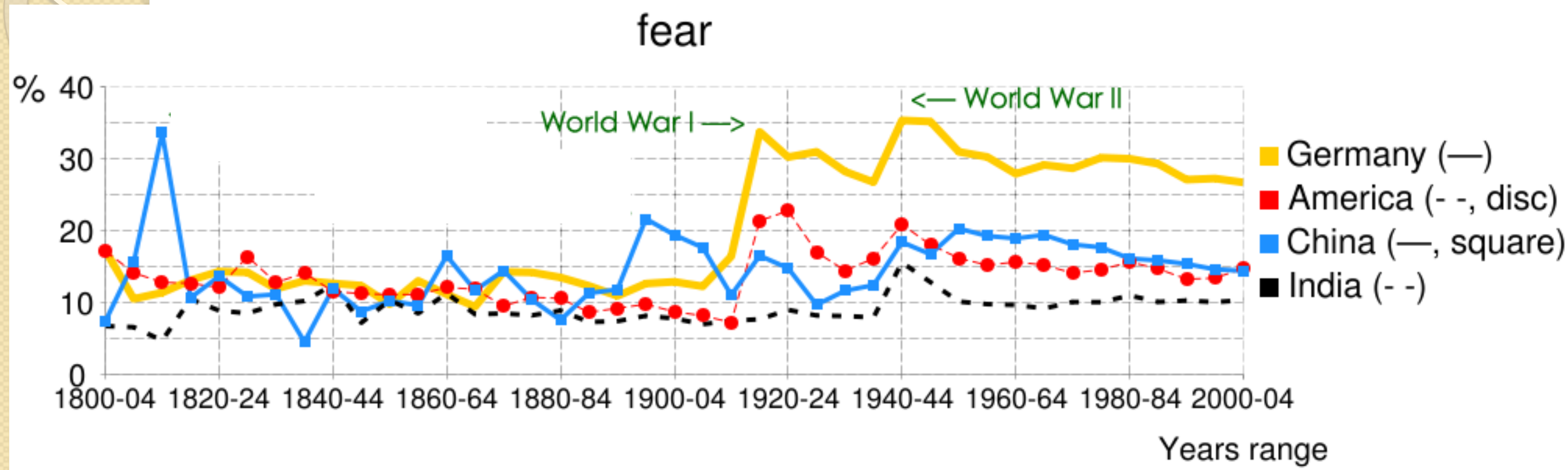
Emotion Word Density

average number of emotion words in every X words

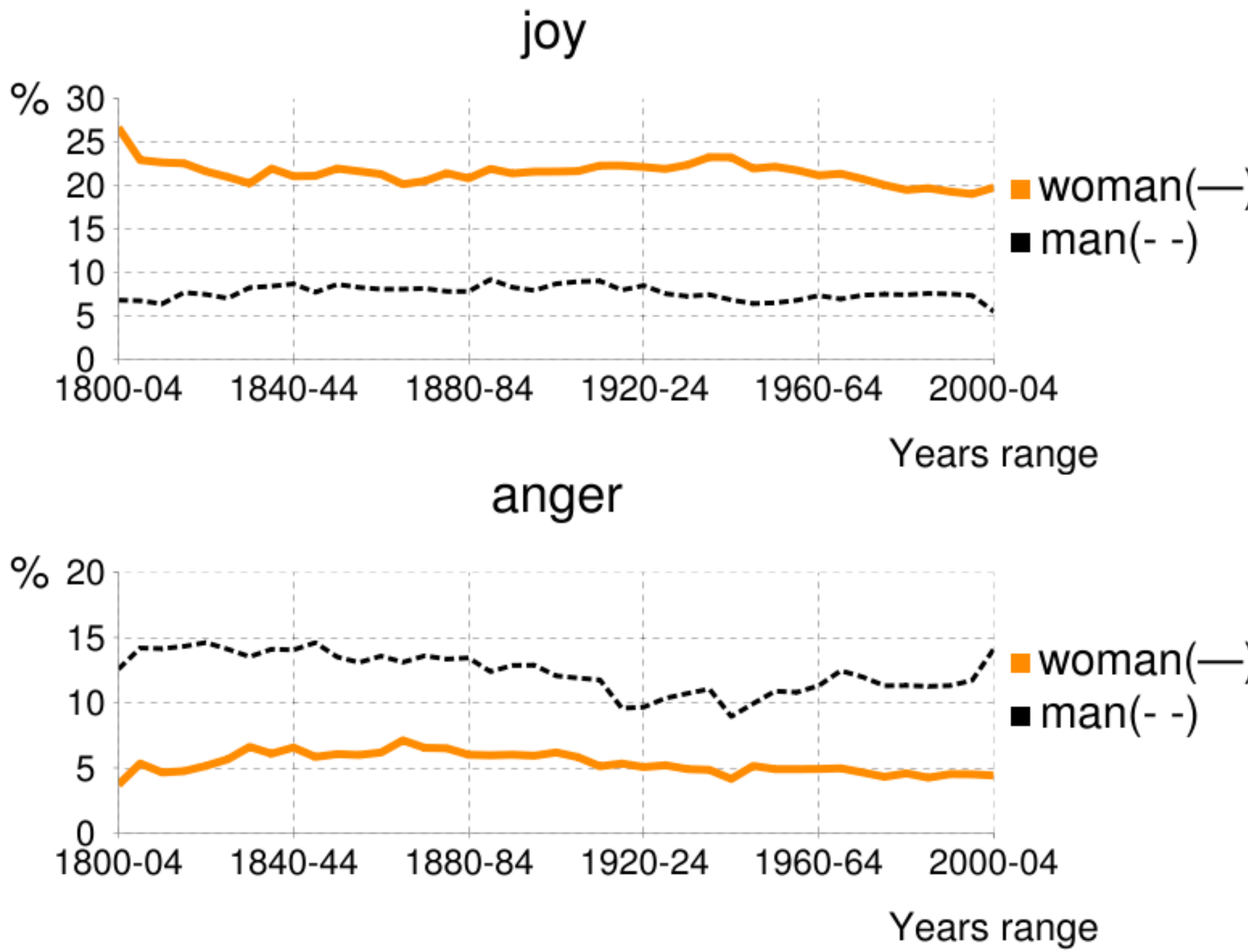


Brothers Grimm fairy tales ordered as per increasing negative word density. X = 10,000.

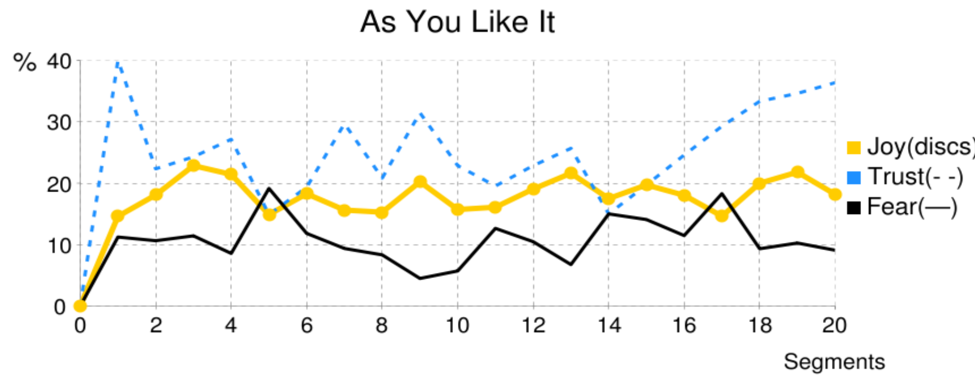
Analysis of Emotion Words in Books

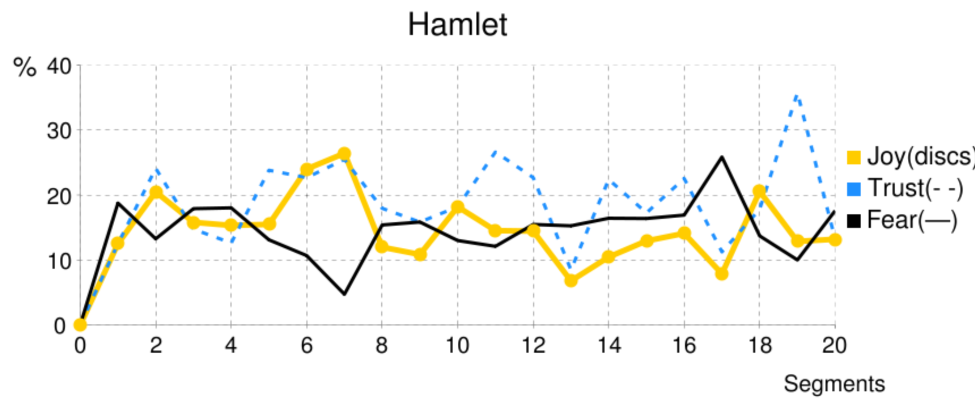


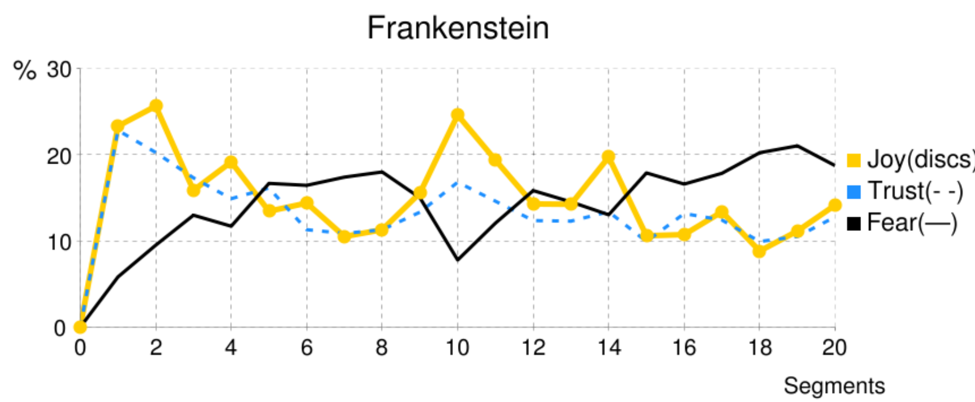
Percentage of fear words in close proximity to occurrences of America, China, Germany, and India in books.



Percentage of joy and anger words in close proximity to occurrences of man and woman in books.







National Research Council Canada

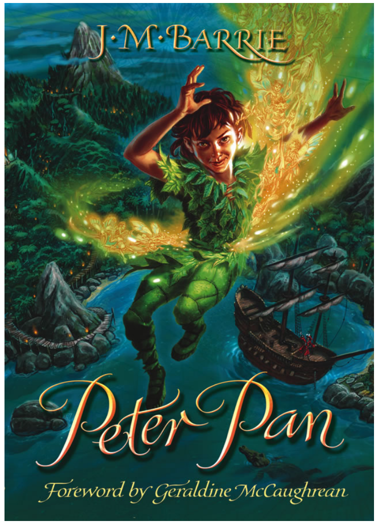
Hannah Davis
Artist/Programmer

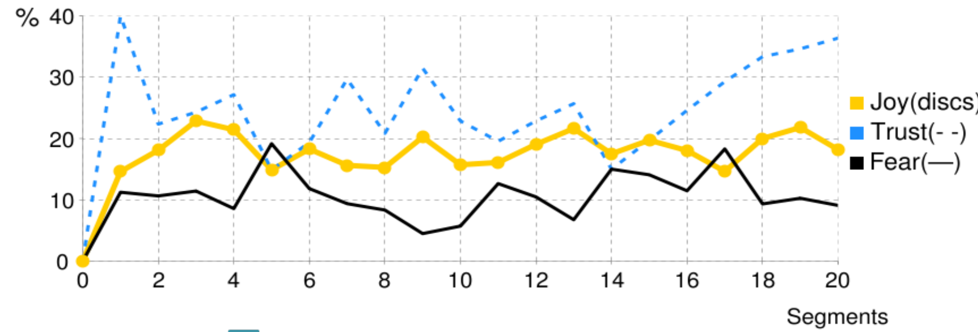
with the sound of

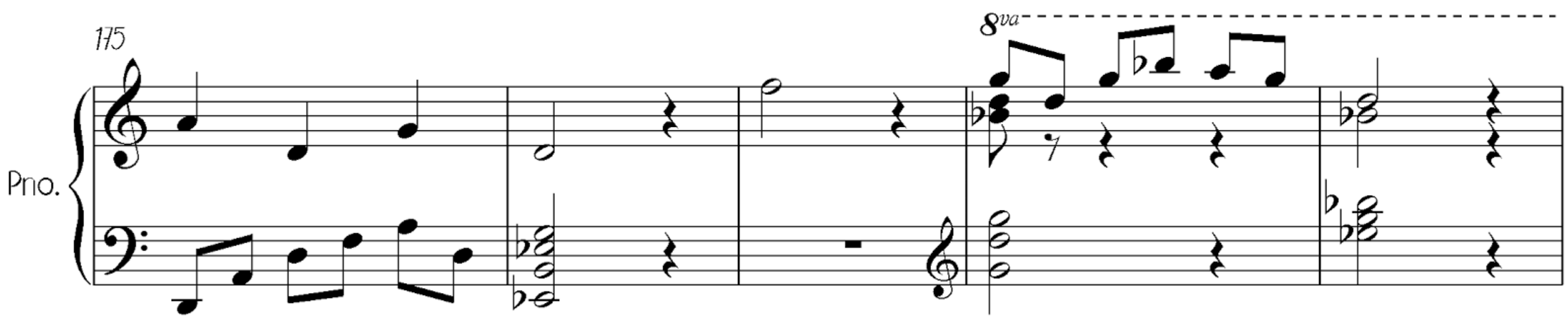
The Words are Alive: ~~Associations with~~ Sentiment, Em☺tion, Colour, and Music ♪

Paper:

- 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.

National Research Council Canada

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

TransProse

- Three simultaneous piano melodies pertaining to the dominant emotions.
- Overall positiveness (or, negativeness) determines:
  - whether C major or C minor
  - base octave
- Partition the novel into small sections
- For each section, if emotion density is high:
  - split section into many small sub-sections and play many short notes corresponding to each sub-section
  - higher emotion densities of a subsection lead to more dissonant notes

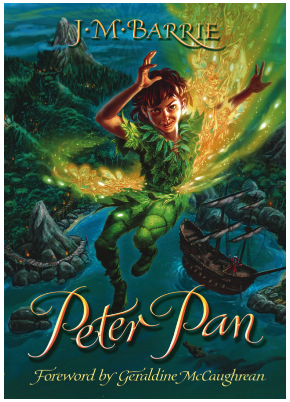
C major: C, D, E, F, G, A, B

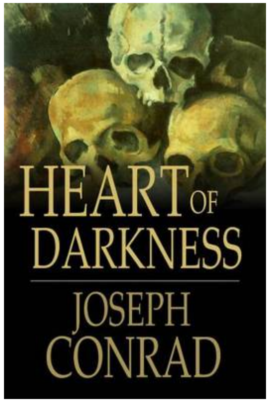
Order of increasing dissonance: C, G, E, A, D, F, B

Pieces

Three simultaneous piano melodies pertaining to the dominant emotions.

Examples





TransProse: www.musicfromtext.com
Music played 300,000 times since website launched in April 2014.

The Words are Alive: Associations with Sentiment, Em☺tion, Colour, and Music

Papers:

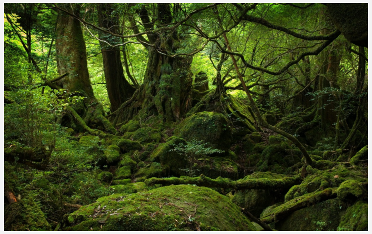
- Colourful Language: Measuring Word-Colour Associations, Saif Mohammad, In Proceedings of the ACL 2011 Workshop on Cognitive Modeling and Computational Linguistics (CMCL), June 2011, Portland, OR.
- Even the Abstract have Colour: Consensus in Word-Colour Associations, Saif Mohammad, In Proceedings of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

Word-Colour Associations

Concrete concepts

→ white

iceberg



→ green

vegetation

Abstract concepts

→ red

danger

→ white

honesty

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

Key Questions

- How much do we agree on word-colour associations?
- How many terms have strong colour associations?
- Can we create a lexicon of such word-colour associations?
- Do concrete concepts have a higher tendency to have colour association?
- How much do colour associations manifest in text?

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	#FBCEB1	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%

Crowdsourced Questionnaire

- 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

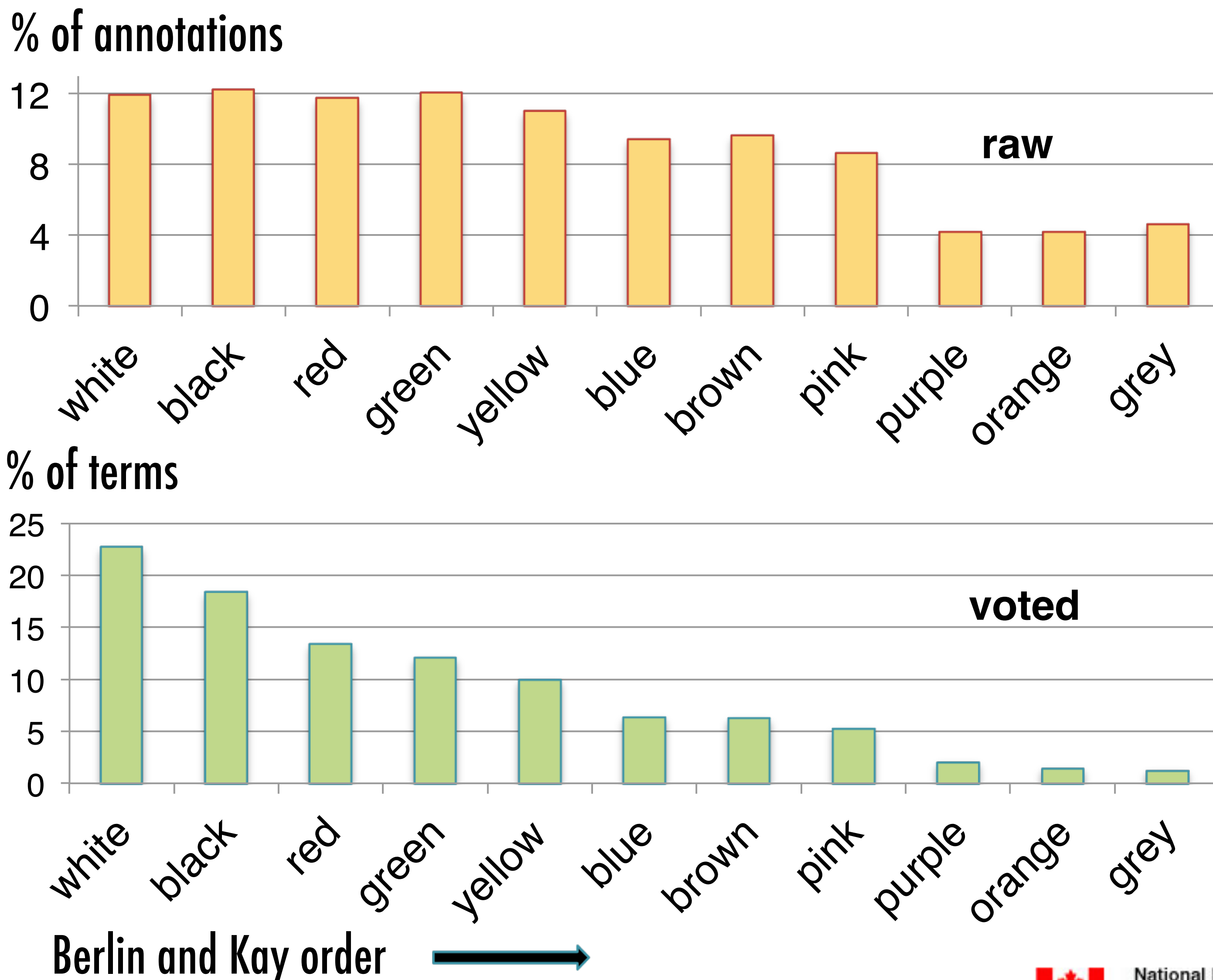
Q. Which colour is associated with *sleep*?

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

NRC Word-Colour Association Lexicon

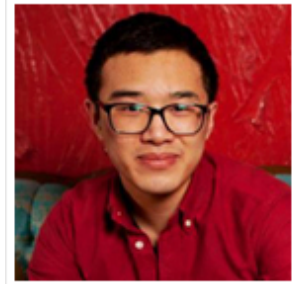
- sense-level lexicon: 24,200 word sense pairs

Associations with Colours



Visualizing Word-Colour Associations

[Visualization](#)




Chris Kim and Chris Collins, UOIT

‹ all words associated with green

PALETTE | A 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 | 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

Key Questions

- How much do we agree on colour associations?
  - About 30% of the terms have strong colour associations.
- Do concrete concepts have a higher tendency to have colour association?
  - Only slightly.
- How much do colour associations manifest in text?
  - Automatic methods obtain accuracies up to 60% (most-frequent class baseline: 33.3%)

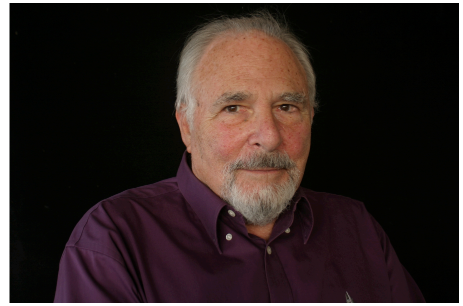
Colourful Language: Measuring Word-Colour Associations, Saif Mohammad, In Proceedings of the ACL 2011 Workshop on Cognitive Modeling and Computational Linguistics (CMCL), June 2011, Portland, OR.

The Words are Alive: Associations with Sentiment, Em☺tion, Colour, and Music

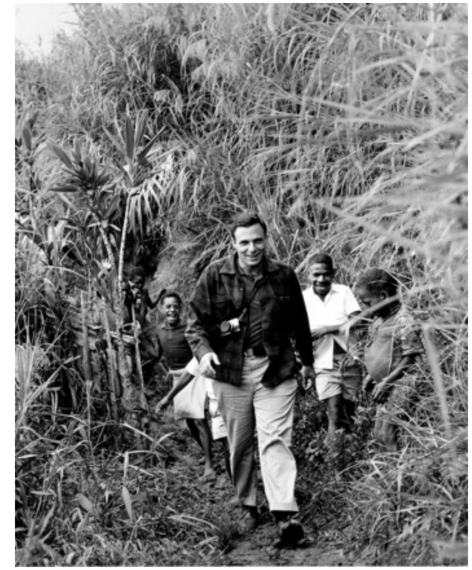
National Research Council Canada

Debate: Universality of Perception of Emotions

Margaret Mead
Cultural anthropologist



Paul Ekman
Psychologist and discoverer of micro expressions.



Lisa Barrett
University Distinguished Professor of Psychology, Northeastern University

- Grad school experiment on people’s ability to distinguish photos of depression from anxiety
  - one is based on sadness, and the other on fear
  - found agreement to be poor
- Agreement also drops for Ekman emotions when participants are given:
  - Just the pictures (no emotion word options)
  - Or say, two scowling faces and asked if the two are feeling the same emotion

No such thing as Basic Emotions?

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

Paper:

#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.

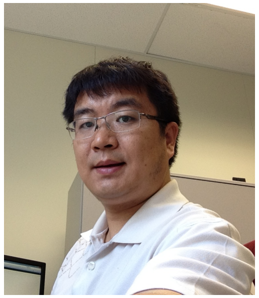
The Words are Alive: Associations with Sentiment, Em☺tion, Colour, and Music ♪

Papers:

- Sentiment Analysis of Short Informal Texts. Svetlana Kiritchenko, Xiaodan Zhu and Saif Mohammad. Journal of Artificial Intelligence Research, 50, August 2014.
- 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.

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

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

National Research Council Canada

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 word w is positive

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

Features of the Twitter Lexicon

- connotation and not necessarily denotation
  - tears, party, vacation
- large vocabulary
  - cover wide variety of topics
  - lots of informal words
  - twitter-specific words
    - creative spellings, hashtags, conjoined words
- seed hashtags have varying effectiveness
  - study on sentiment predictability of different hashtags (Kunneman, F.A., Liebrecht, C.C., van den Bosch, A.P.J., 2014)

Negation

Jack was not thrilled at the prospect of working weekends ☹

negator

sentiment
label: negative

need to determine this word's
sentiment when negated

The bill is not garbage, but we need a more focused effort ☹

negator

sentiment
label: negative

need to determine this word's
sentiment when negated

Handling Negation

Jack was not thrilled at the prospect of working weekends ☹

in the list of negators

scope of negation

sentiment label: negative

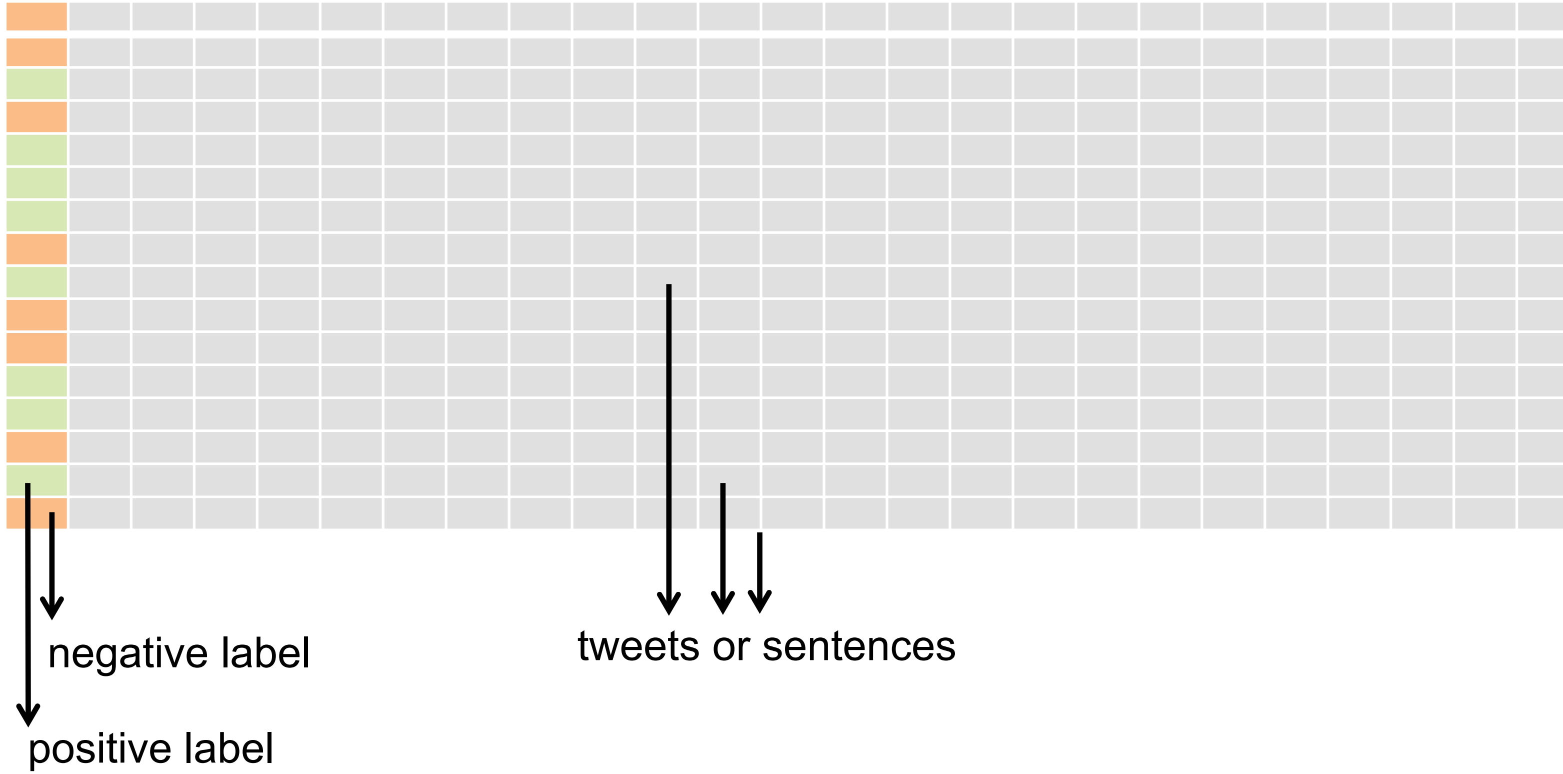
The bill is not garbage, but we need a more focused effort ☹

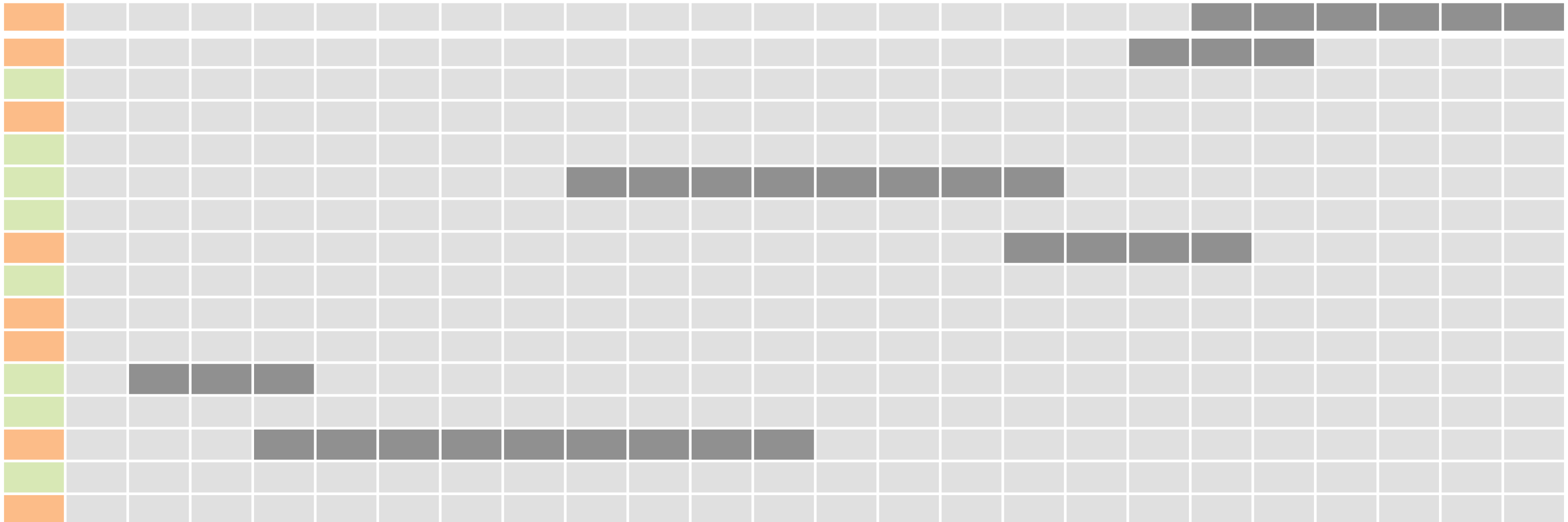
in the list of negators

scope of negation

sentiment label: negative

Scope of negation: from negator till a punctuation (or end of sentence)


negative label
tweets or sentences
positive label



affirmative contexts
(in light grey)

negated contexts
(in dark grey)

Generate sentiment lexicon for words in affirmative context

Generate sentiment lexicon for words in negated context

Term	Sentiment140 Lexicons		
	Base	**AffLex**	**NegLex**
Positive terms			
great	1.177	1.273	-0.367
nice	0.974	1.149	-0.912
honest	0.391	0.431	-0.123
Negative terms			
terrible	-1.766	-1.850	-0.890
bad	-1.297	-1.674	0.021
negative	-0.090	-0.261	0.389

Table 3: Example sentiment scores from the Sentiment140 Base, Affirmative Context (AffLex) and Negated Context (NegLex) Lexicons.

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

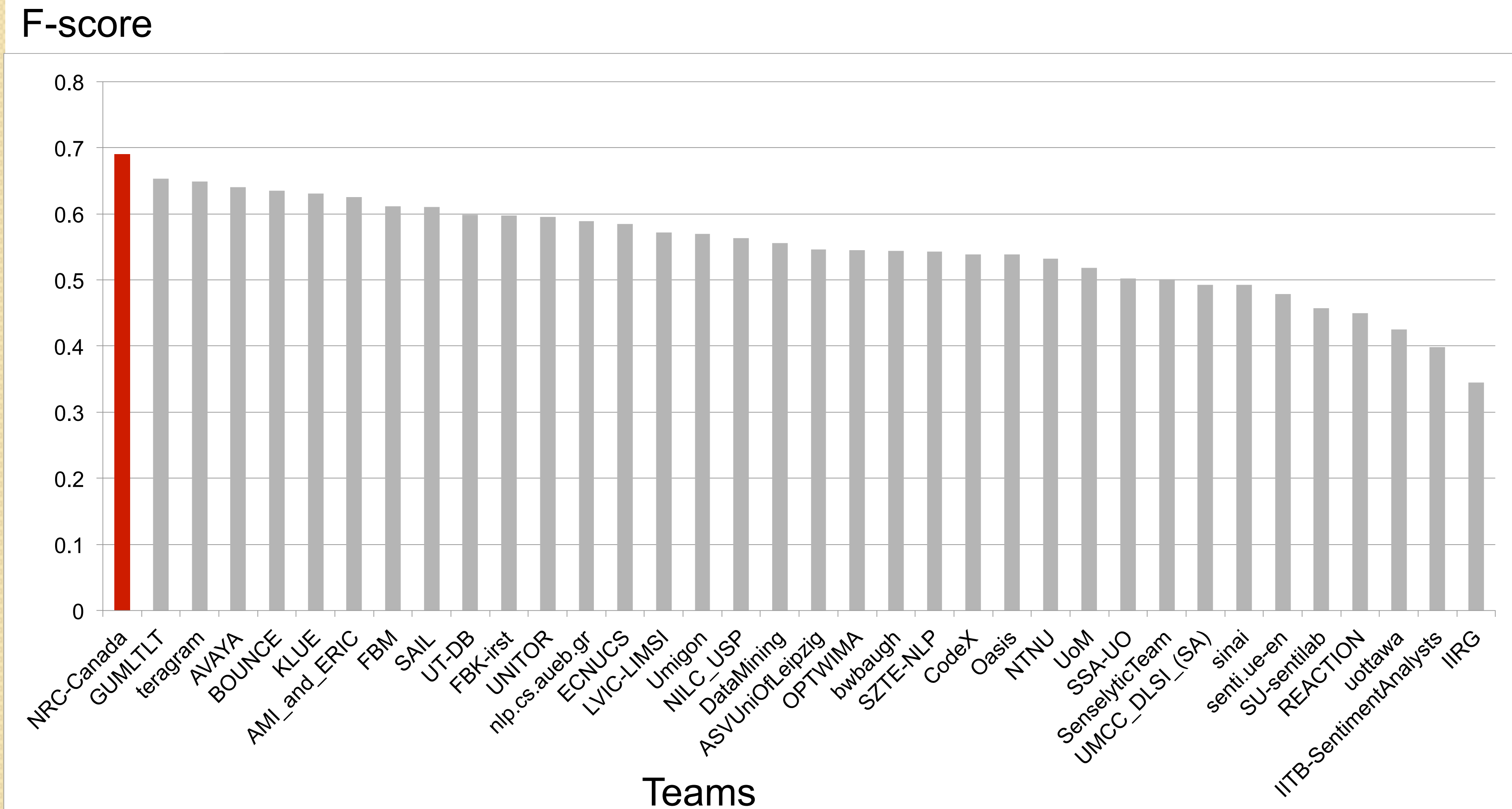
Features	Examples
sentiment lexicon	#positive: 3, scorePositive: 2.2; maxPositive: 1.3; last: 0.6, scoreNegative: 0.8, scorePositive_neg: 0.4
word n-grams	spectacular, like documentary
char n-grams	spect, docu, visua
part of speech	#N: 5, #V: 2, #A:1
negation	#Neg: 1; ngram:perfect → ngram:perfect_neg, polarity:positive → polarity:positive_neg
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: 1^{st} rank
    - term level: 1^{st} rank
  - SMS messages
    - message-level: 1^{st} rank
    - term level: 2^{nd} rank

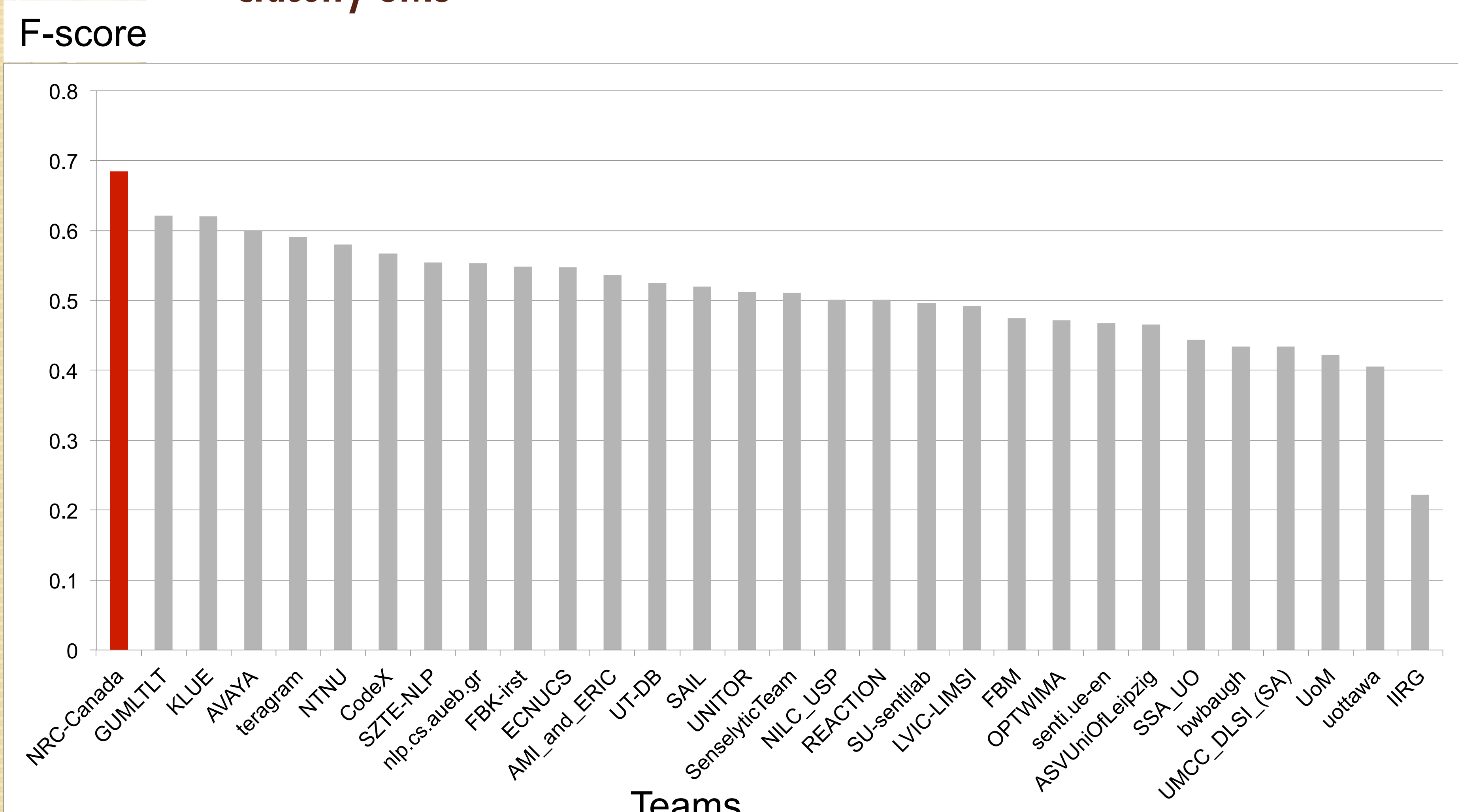
Sentiment Analysis Competition

Classify Tweets



National Research Council Canada

Sentiment Analysis Competition
Classify SMS



National Research Council Canada

NRC-Canada in SemEval-2013, Task 2

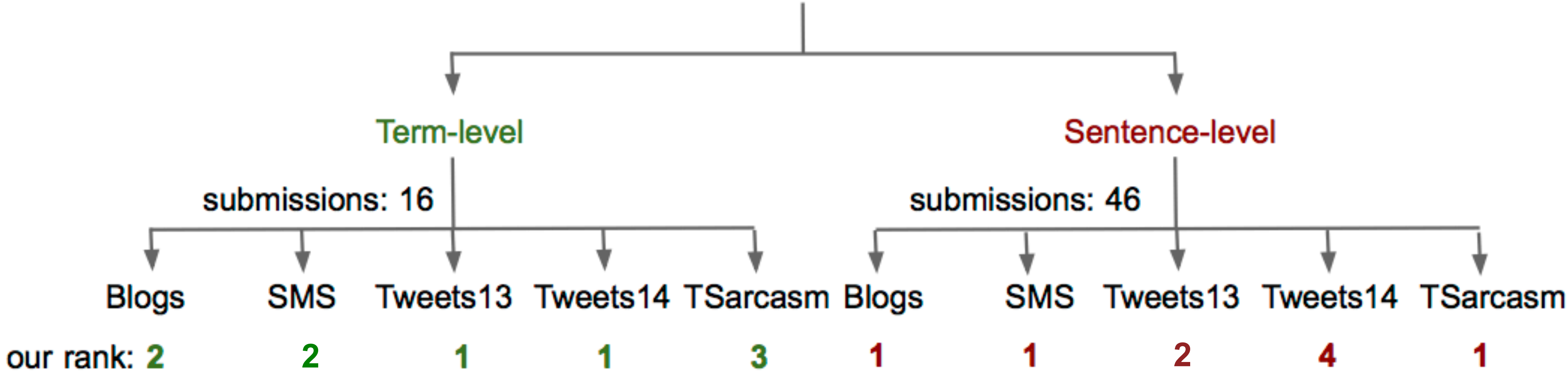
Released description of features.
Released resources created (tweet-specific sentiment lexicons).
www.purl.com/net/sentimentoftweets

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.

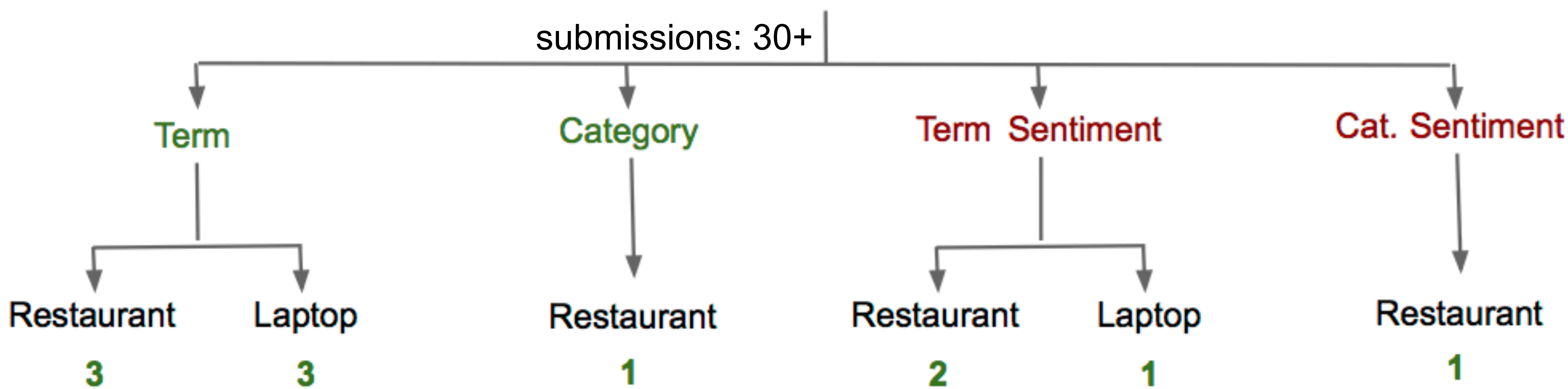
Sentiment Analysis of Social Media Texts (SemEval-2014 Task 9)

	Term-level					Sentence-level				
	submissions: 16					submissions: 46				
	Blogs	SMS	Tweets13	Tweets14	TSarcasm	Blogs	SMS	Tweets13	Tweets14	TSarcasm
our rank:	2	2	1	1	3	1	1	2	4	1

Sentiment Analysis of Social Media Texts (SemEval-2014 Task 9)



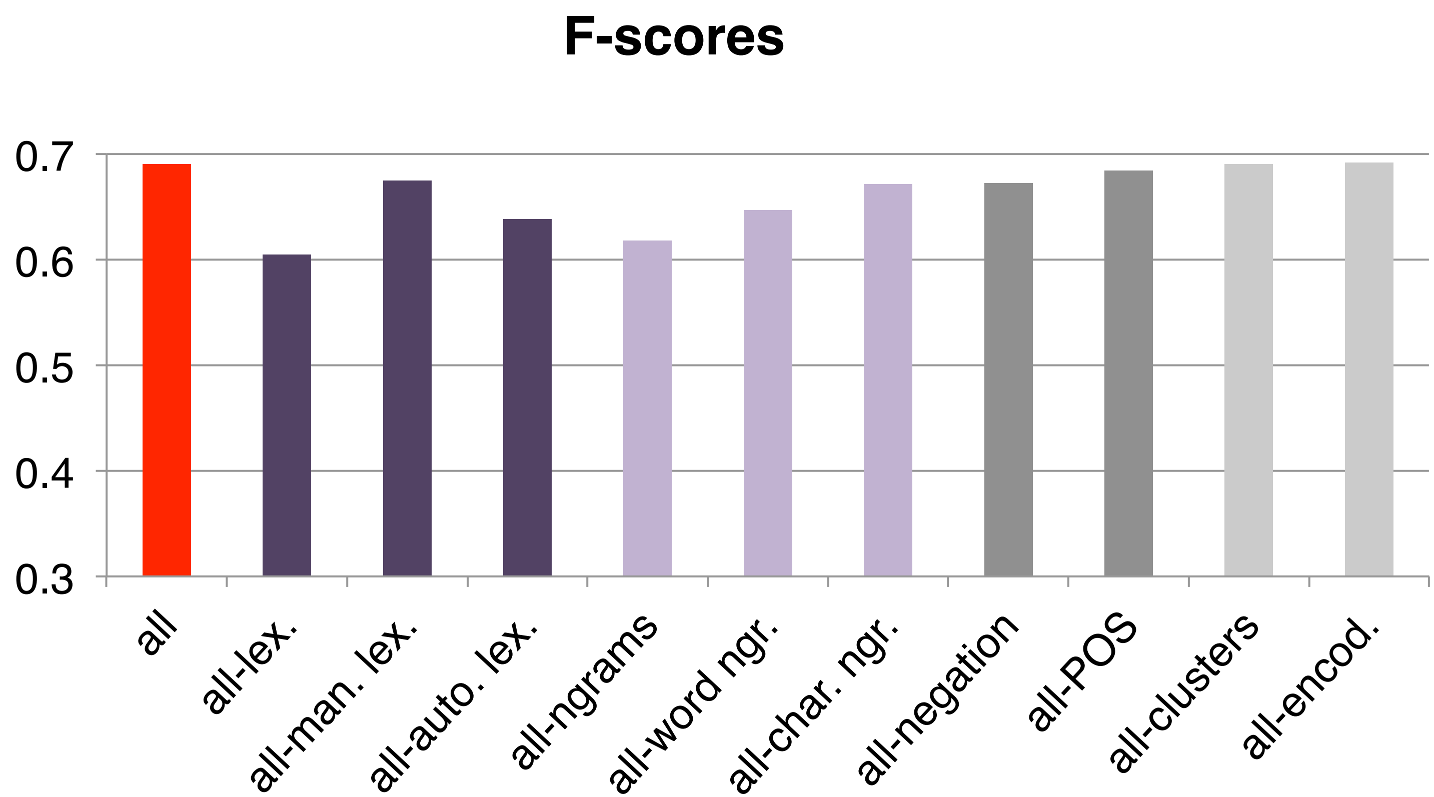
Aspect-Based Sentiment Analysis (SemEval-2014 Task 4)



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)



National Research Council Canada

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**

How to manually create sentiment lexicons with intensity values?

- Humans are not good at giving real-valued scores?
  - hard to be consistent across multiple annotations
  - difficult to maintain consistency across annotators
    - 0.8 for annotator may be 0.7 for another
- Humans are much better at comparisons
  - Questions such as: Is one word more positive than another?
  - Large number of annotations needed.

Need a method that preserves the comparison aspect, without greatly increasing the number of annotations needed.

MaxDiff

- The annotator is presented with four words (say, A, B, C, and D) and asked:
  - which word is the most positive
  - which is the least positive

- By answering just these two questions, five out of the six inequalities are known
  - For e.g.:
    - If A is most positive
    - and D is least positive, then we know:

$$A > B, A > C, A > D, B > D, C > D$$

MaxDiff

- Each of these MaxDiff questions can be presented to multiple annotators.
- The responses to the MaxDiff questions can then be easily translated into:
  - a ranking of all the terms
  - a real-valued score for all the terms (Orme, 2009)
- If two words have very different degrees of association (for example, A >> D), then:
  - A will be chosen as most positive much more often than D
  - D will be chosen as least positive much more often than A.

  This will eventually lead to a ranked list such that A and D are significantly farther apart, and their real-valued association scores will also be significantly different.

Dataset of Sentiment Scores
(Kiritchenko, Zhu, and Mohammad 2014)

- Selected ~1,500 terms from tweets
  - regular English words: peace, jumpy
  - tweet-specific terms
    - hashtags and conjoined words: #inspiring, #happytweet, #needsleep
    - creative spellings: amazzing, goooood
  - negated terms: not nice, nothing better, not sad

- Generated 3,000 MaxDiff questions
- Each question annotated by 10 annotators on CrowdFlower
- Answers converted to real-valued scores (0 to 1) and to a full ranking of terms using the counting procedure (Orme, 2009)

Examples of sentiment scores from the MaxDiff annotations

Term	Sentiment Score 0 (most negative) to 1 (most positive)
awesomeness	0.9133
#happygirl	0.8125
cant waitttt	0.8000
don't worry	0.5750
not true	0.3871
cold	0.2750
#getagrip	0.2063
#sickening	0.1389

Robustness of the Annotations

- Divided the MaxDiff responses into two equal halves
- Generated scores and ranking based on each set individually
- The two sets produced very similar results:
  - average difference in scores was 0.04
  - Spearman's Rank Coefficient between the two rankings was 0.97

Dataset will be used as test set for Subtask E in Task 10 of SemEval-2015: Determining prior probability.

Trial data already available:

http://alt.qcri.org/semeval2015/task10/index.php?id=data-and-tools
(Full dataset to be released after the shared task competition in Dec., 2014.)

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

Mechanical Turk Annotations

- NRC word-emotion and word-sentiment 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
- MaxDiff word-sentiment association Lexicon (with intensity scores)
  - Entries for 15,000 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

www.saifmohammad.com
saif.mohammad@nrc-cnrc.gc.ca