



From Once Upon a Time to Happily Ever After: Tracking Emotions in Books and Mail

Saif Mohammad

National Research Council Canada



Road Map

- Motivation and background
- Emotion Lexicon
- Analysis of emotion words in mail
- Analysis of emotion words in books

Source papers:

[Crowdsourcing a Word-Emotion Association Lexicon](#)

Saif Mohammad and Peter Turney, To Appear In Computational Intelligence, Wiley Blackwell Publishing Ltd.

[Tracking Sentiment in Mail: How Genders Differ on Emotional Axes](#)

Saif Mohammad and Tony Yang, In Proceedings of the 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.



Road Map

- Motivation and background
- **Emotion Lexicon**
- Analysis of emotion words in mail
- Analysis of emotion words in books

Source papers:

Crowdsourcing a Word-Emotion Association Lexicon

Saif Mohammad and Peter Turney, To Appear In Computational Intelligence, Wiley Blackwell Publishing Ltd.

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.

Road Map

- Motivation and background
- Emotion Lexicon
- **Analysis of emotion words in mail**
- Analysis of emotion words in books

Source papers:

Crowdsourcing a Word-Emotion Association Lexicon

Saif Mohammad and Peter Turney, To Appear In Computational Intelligence, Wiley Blackwell Publishing Ltd.

Tracking Sentiment in Mail: How Genders Differ on Emotional Axes

Saif Mohammad and Tony Yang, In Proceedings of the 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.



Road Map

- Motivation and background
- Emotion Lexicon
- Analysis of emotion words in mail
- **Analysis of emotion words in books**

Source papers:

Crowdsourcing a Word-Emotion Association Lexicon

Saif Mohammad and Peter Turney, To Appear In Computational Intelligence, Wiley Blackwell Publishing Ltd.

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.

Acknowledgments



Peter Turney
National Research Council Canada



Tony (Wenda) Yang
Simon Fraser University



Emotions



(Phil, from the San Francisco Chronicle)
speaker/writer



When your cartoon can get you killed



listener/reader



Death threats over South Park episode
Event



Extremists
Participants



Trey Parker, Matt Stone
Participants

Words



associated with joy

When your **cartoon** can get you **killed**



associated with sadness

Our goal

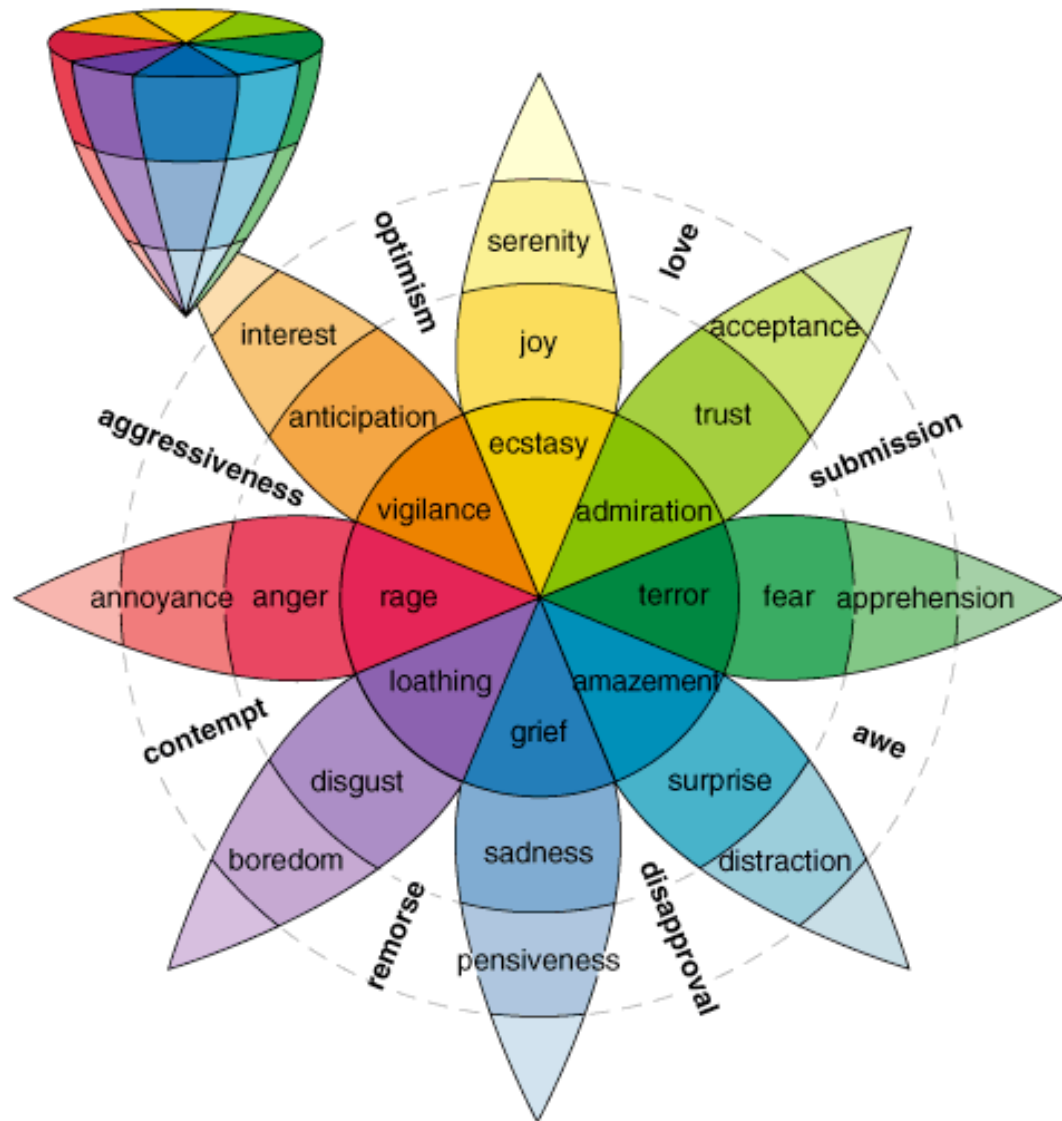
- Create a large word-emotion association lexicon through input from people.
 - Examples:
 - **vampire** is typically associated with fear
 - **startle** is associated with surprise
 - **bliss** is associated with joy
 - **death** is associated with sadness
 - **eager** is associated with anticipation
- Use the lexicon to understand the use of emotion words in text.

Which Emotions?



Plutchik, 1980: Eight Basic Emotions

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



Using Mechanical Turk for



**CROWDSOURCING A
WORD-EMOTION ASSOCIATION
LEXICON**



Crowdsourcing

- **Benefits**
 - Inexpensive
 - Convenient and time-saving
 - Especially for large-scale annotation
- **Challenges**
 - Quality control
 - Malicious annotations
 - Inadvertent errors



Target n-grams

Must be:

- in *Roget's Thesaurus*
- high-frequency term in the Google n-gram corpus

Word-Choice Question

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

• *car* • *tree* • *fish* • *olive*

- Generated automatically
 - Near-synonym taken from thesaurus
 - Distractors are randomly chosen
- Guides Turkers to desired sense
- Aides quality control
 - If Q1 is answered incorrectly:
 - Response to Q2 is discarded



Association Questions

Q2. How much is *shark* associated with the emotion fear?

(for example, *horror* and *scary* are strongly associated with fear)

- *shark* is not associated with fear
 - *shark* is weakly associated with fear
 - *shark* is moderately associated with fear
 - *shark* is strongly associated with fear
-
- Eight such questions for the eight emotions.
 - Two such questions for positive or negative.

Emotion Lexicon

- Each word-sense pair is annotated by 5 Turkers
- About 10% of the assignments were discarded due to incorrect response to Q1 (gold question)
- Targets with less than 3 valid assignments removed
- **NRC Emotion Lexicon**
 - **sense-level lexicon**
 - word sense pairs: 24,200
 - **word-level lexicon**
 - union of emotions associated with the different senses of a word
 - word types: 14,200



**MOTIVATION:
EMOTION ANALYSIS OF MAIL**



Applications of emotion analysis of mail

- Tracking emotions in one's own email
- Developing a depression—happiness index
 - Example: determining risk of repeat suicide attempt (Osgood and Walker, 1959; Pestian et al., 2008)
- Search
 - Example: locating snippets expressing anger, for customer relation management (Diaz and Ruz, 2002; Dube and Maute, 1996)
- Analyzing Social Behaviour
 - Example: how genders communicate in the work place
- Analyzing Social Networks
 - Example: Tracking flow of emotions in a social network
- Developing Writing Aids
 - Example: Aids for avoiding misinterpretation (Liu et al., 2003)



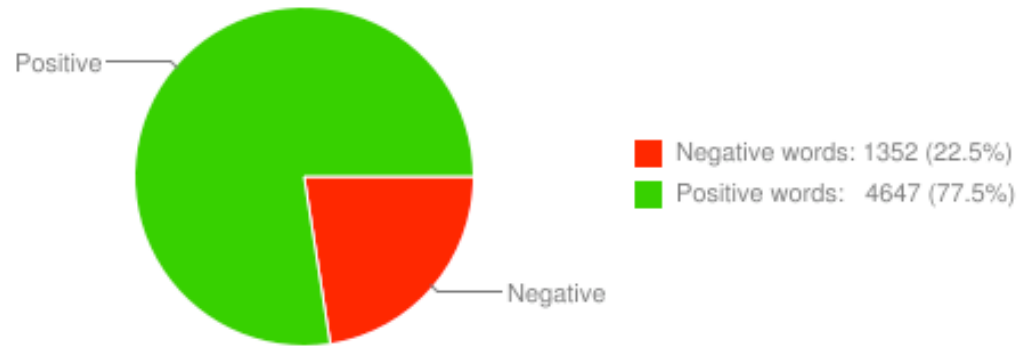
Sources of Mail for Research

- **Enron Email Corpus** (Klimt and Yang, 2004)
 - Work-place communication
 - 200,000 emails
 - October 1998 to June 2002
 - 150 of the senior management at Enron Corporation
- **ECUE Spam Datasets**
- **Phishing Datasets**

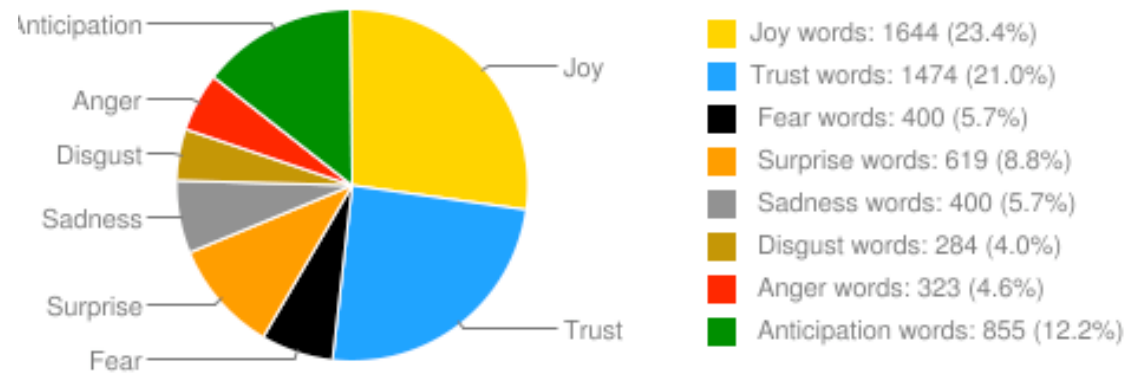
Sources of Mail for Research (continued)

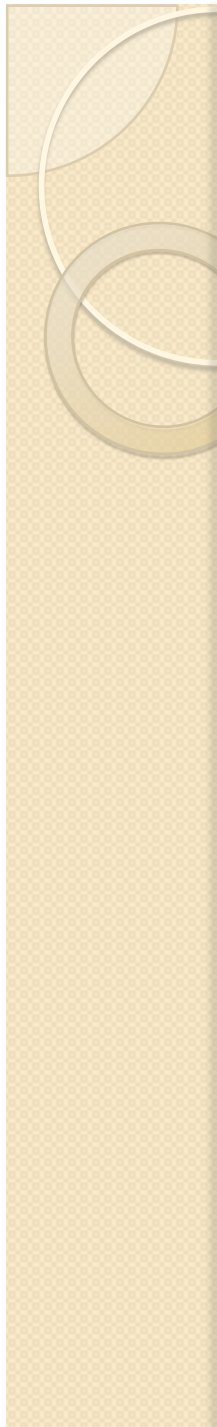
- Suicide Notes Corpus
 - 21 notes posted by Art Kleiner
 - I2B2 2011 will release a few thousand suicide notes
- Hate Mail Corpus
 - 279 pieces of hate mail sent to the *Millenium Project*
- Love Letters Corpus
 - 348 postings at lovingyou.com

love letters

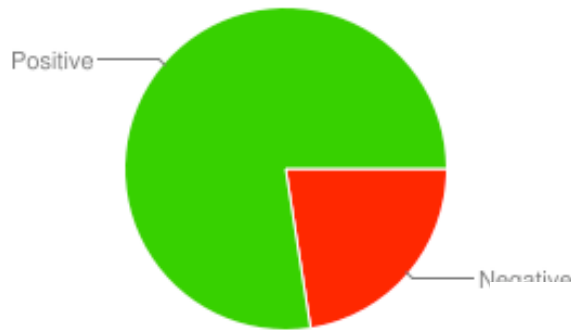


love letters



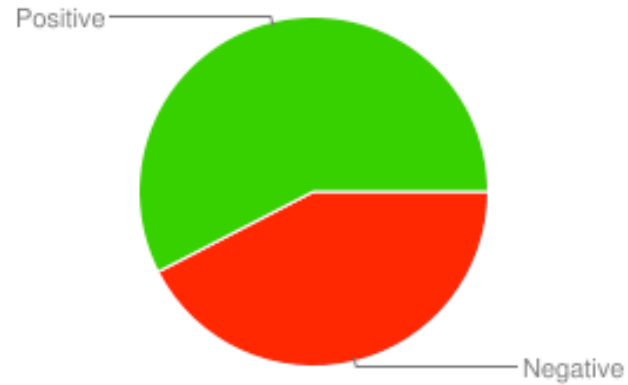


love letters



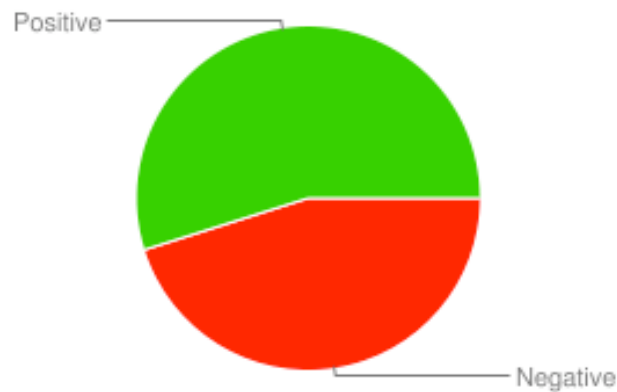
■ Negative words: 1352 (22.5%)
■ Positive words: 4647 (77.5%)

hate mail

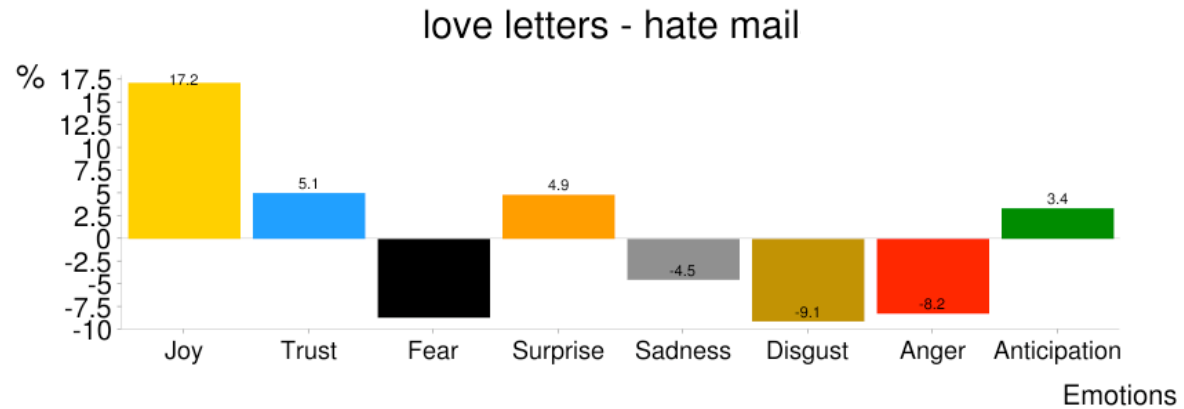
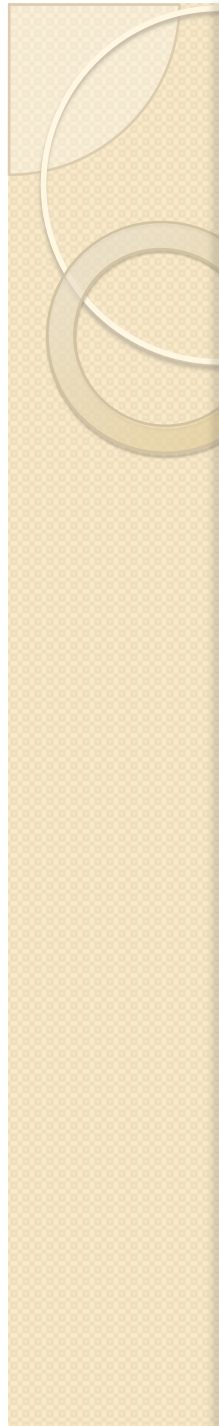


■ Negative words: 2999 (42.4%)
■ Positive words: 4071 (57.6%)

suicide notes

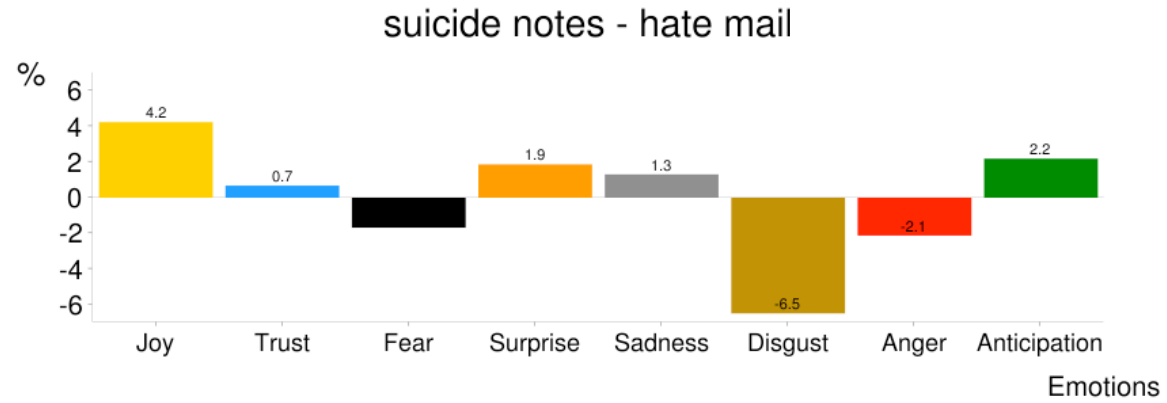
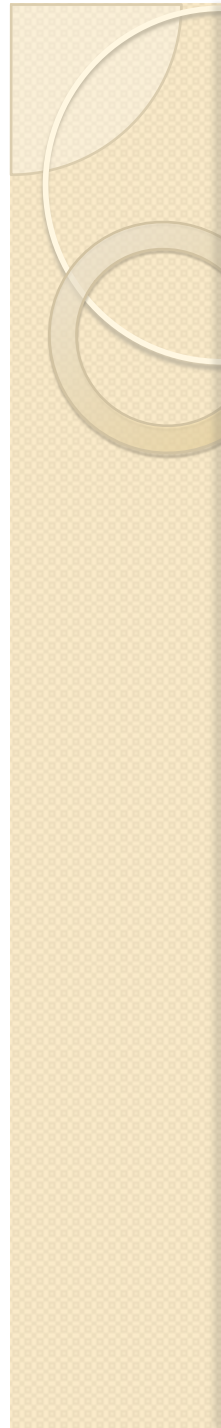


■ Negative words: 228 (45.1%)
■ Positive words: 277 (54.9%)



lucky beautiful feeling wonderful
hope smile perfect loving thankful
blessed baby true grow kiss glad treasure
marriage gift promise dove happiness joy laugh music
undying cream sweetheart gorgeous inspiration intelligence

relative salience of joy words



horrible powerful lying cancer dying
excrement ignorant quack loon blame
sick hatred fraudulent rat flea john murder finally
lesser harass denounce slander imprisoned illegal garbage
disease adverse crap tetanus disaster
contamination epidemic toxic violent threatening vacuous
lose arrogant displeasure gag damage distress venom

relative salience of disgust words



WOMEN vs. MEN: Comparative Analysis of
EMOTION WORDS IN WORK-PLACE EMAIL

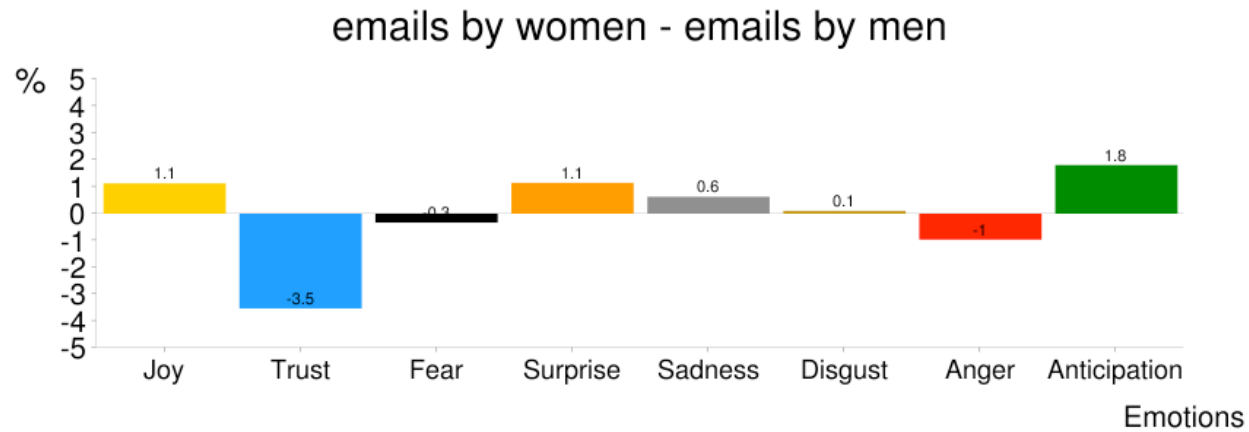
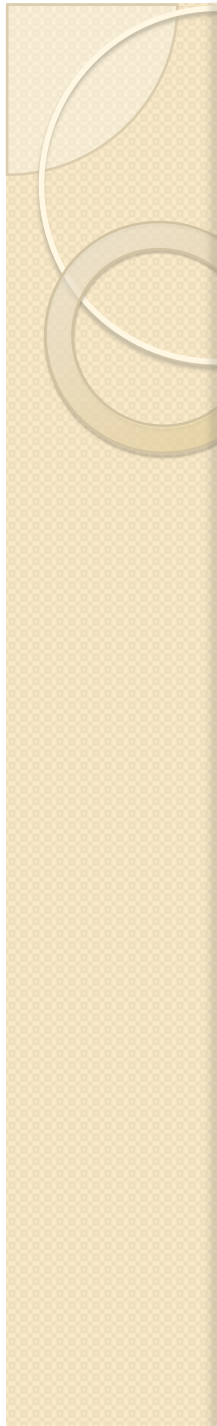
Genders and Language

- Huge literature at the intersection of gender and language (see bibliographies by Schiffman, 2002 and Sunderland et al., 2002)
- Men and women use language differently, even in computer-mediated communication (Boneva et al., 2001)
 - Women
 - foster personal relations (Deaux and Major, 87; Eagly and Steffen, 84)
 - share concerns and support others (Boneva et al., 2001)
 - Men
 - communicate for social position (Tannen, 1991)
 - prefer to talk about activities (Caldwell and Peplau, 1982; Davidson and Duberman, 1982)

What are the differences in how genders use emotion words in work-place email?

Email Corpus

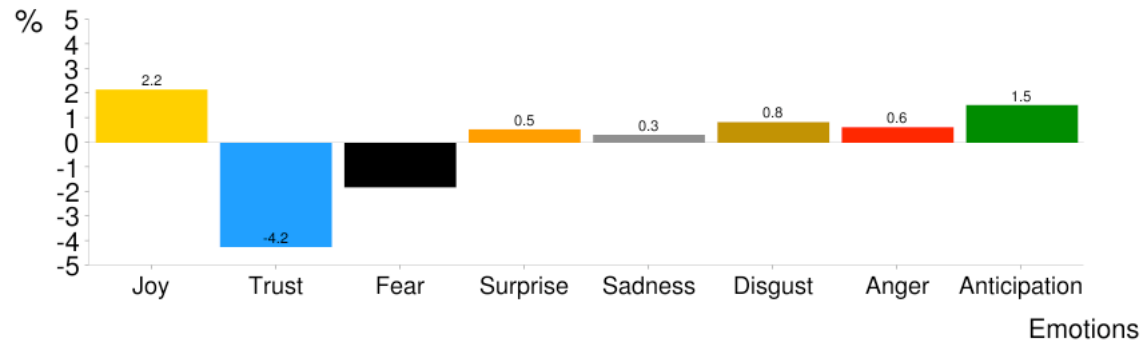
- Used the Enron Email Corpus
- Removed mails with less than 50 and more than 200 words (Cheng et al., 2009)
- Identified gender of senders by name
 - 41 female, 89 male, 20 untagged
- Mails sent by gender-unknown employees removed
- 32,045 mails remaining
 - 19,920 sent by men
 - 12,125 sent by women



prepared legal friend haven kind accurate brother
important feeling excited invite cash found
agreed understanding prefer confirmed verified
true frank top shepherd corporation grow
playground fortune **money** improve teacher
pretty explain real majority repay buddy
volunteer food fundamental tree respect

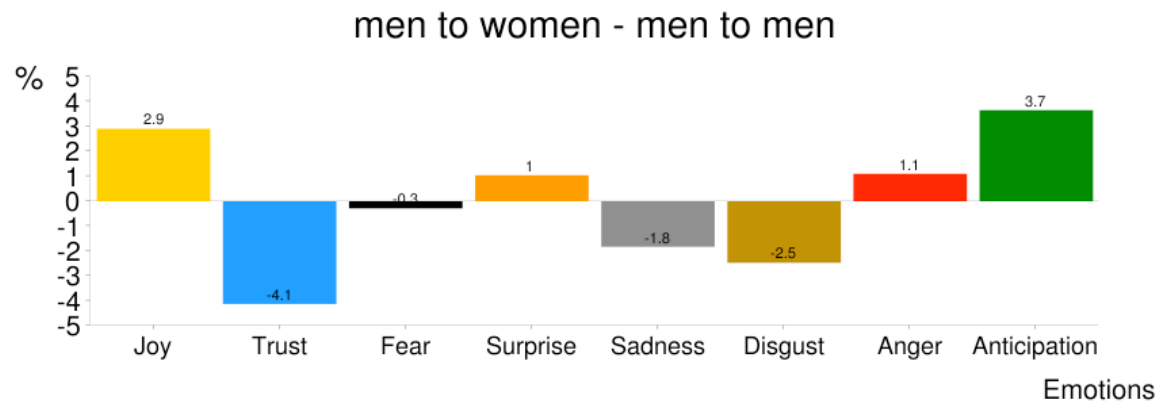
relative salience of trust words

emails to women - emails to men



wonderful enthusiasm satisfied **luck** massage abundance
excellent exceed volunteer evergreen resources ordination
baby gift child favorite **true** buddy tree lover hilarious
perfect clown playground award **success**
clean holiday merry entertainment blessed joy
birth pastor sing spirits gain **diamond electric** kiss
surprise food presto favorable beautiful
garden excited thanksgiving pleasant victory

relative salience of joy words



completion fun feeling white pretty

hope death marry hero ultimately

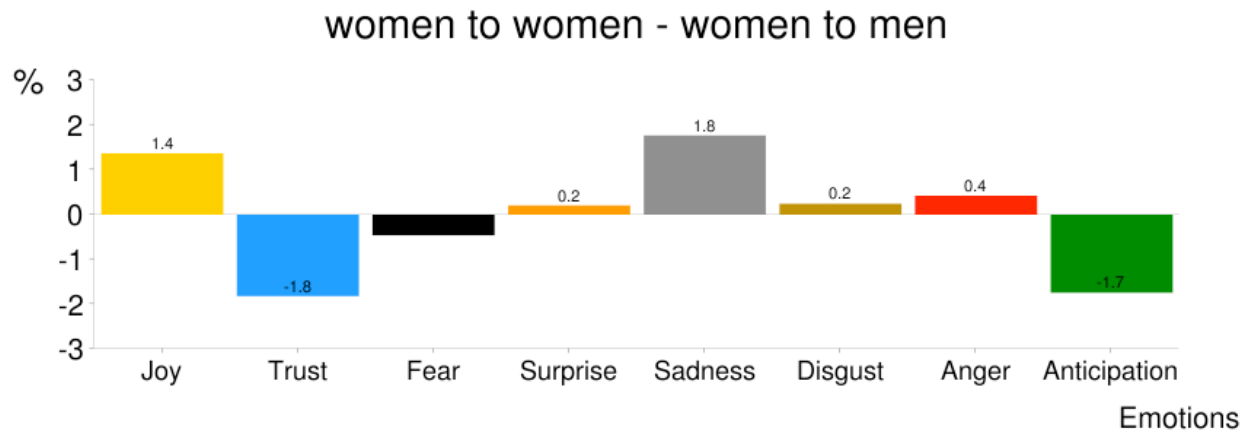
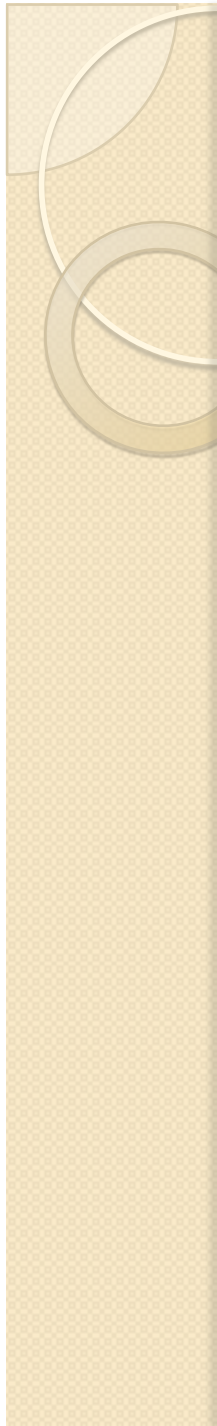
invitation immediately destination intended gift

child luck cautious prepare judicial buddy tree

anxious glad lover clown playground finally enjoy award

birth gain holiday wait nervous discretion sterling precious

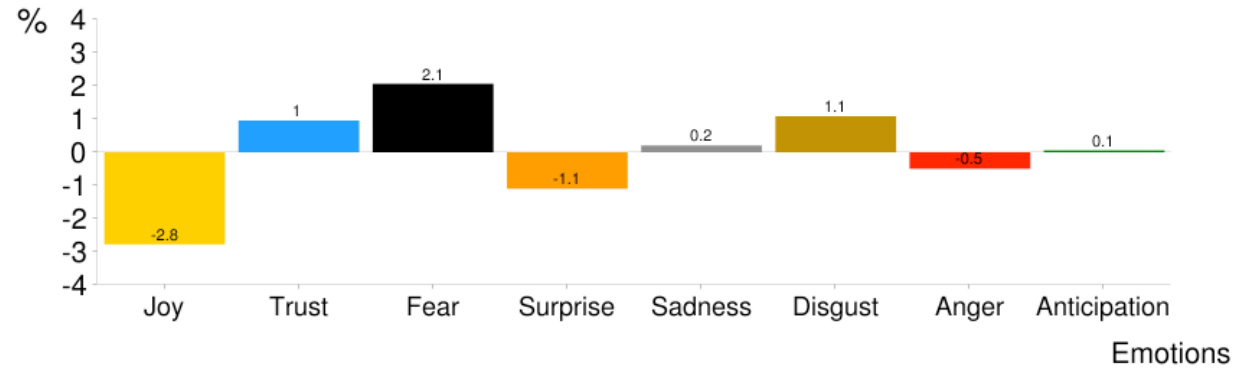
relative salience of anticipation words



late **termination** missing sue sorrow tragedy
punished rob constraint art cancer regret bankruptcy
isolated cross bottom **problem** elimination
unsuccessful **feeling mistake** gore resign bankrupt
wanting sing sick exhausted broke adverse **error**
undesirable cumbersome violently kill wan struggle **damage**
margin ill hell crazy doubt **resentment** mourning
pointless misunderstanding woe **black** cancel bitch
condescension sin meaningless conflict hunter
lynch interrupted avalanche moody **delay** guilty

relative salience of sadness words

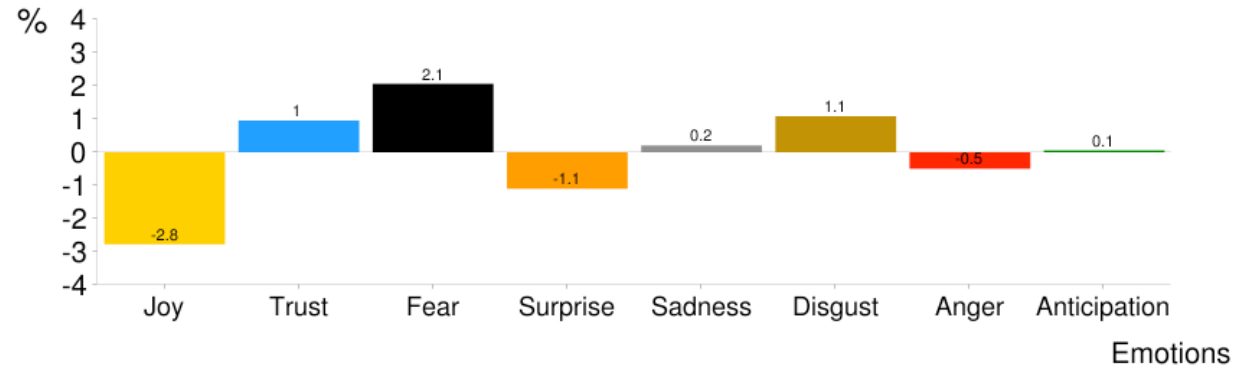
men to men - women to women



evil regulatory risk problem acrobat destination
 prohibited nervous predicament shaky afraid
hearing forced attack lose lunacy avoid
 missing worse attorney reluctant illegitimate horrible
 bankruptcy government rob suspension suspect terminal
 manipulation killing lawsuit hunting gun forgotten discourage
 procedure confusion emergency hunter punitive
 uncertain urgent threat opposed abomination police abandoned

relative salience of fear words

men to men - women to women



hope improvement create resources
electric promise respect holiday glad hire
proud sun presto enjoy feeling joy pleased accomplish
outstanding merry **success** white perfect
accomplished diamond exciting excited luck inviting exceed lucky

relative salience of joy words



MOTIVATION: EMOTION ANALYSIS OF BOOKS



Number of Books Published in a Year

(source: Wikipedia)

United States (2009) 288,355 ("new titles and editions") [3]

United Kingdom (2005) 206,000 [2]

China (2007) 136,226 [4]

Russian Federation (2008) 123,336 [5]

Germany (2009) 93,124 (new titles) [6]

Spain (2008) 86,300 [7]

India (2004) 82,537 (21,370 in Hindi and 18,752 in English) [8][9]

Japan (2009) 78,555 [10]

Iran (2010) 65,000 [11][12]

Taiwan (2007) 42,018 [13]



Sources of Digitized Books

- Project Gutenberg:
 - more than 34,000 books
- Google Books Corpus (GBC):
 - 5.2 million books published from 1600 to 2009
 - English portion has 361 billion words
 - 1-grams, 2-grams, 3-grams, 4-grams, 5-grams



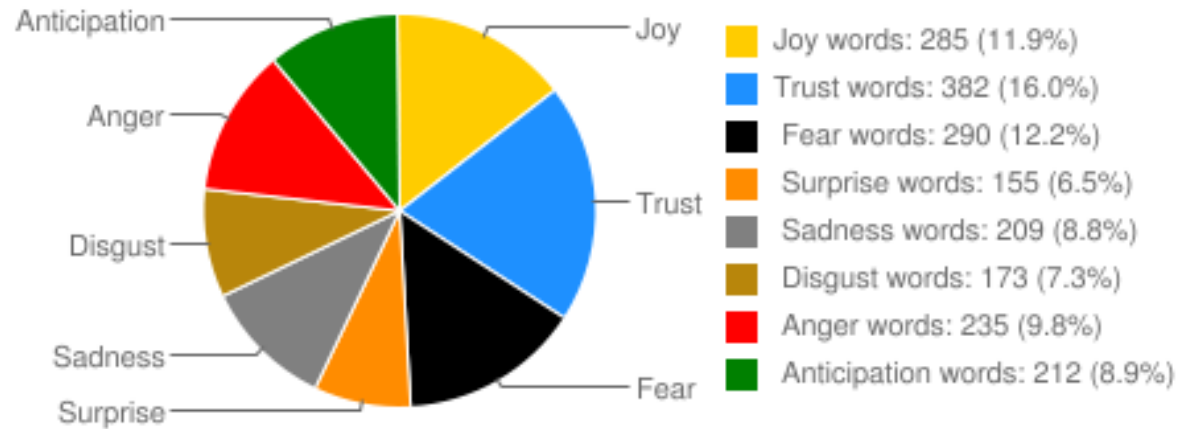
Applications of emotion analysis of books

- Search
 - Example: Which Brothers Grimm tales are the darkest?
- Social Analysis
 - Example: How have books portrayed entities over time?
(Michel et al. 2011)
- Literary Analysis
 - Example: Is the distribution of emotion words in fairy tales significantly different from that in novels?
- Summarization
 - Example: Automatically generate summaries that capture different emotional states of characters in a novel
- Analyzing Persuasion Tactics
 - Example: how emotion words are used for persuasion?
(Mannix, 1992; Bales, 1997)

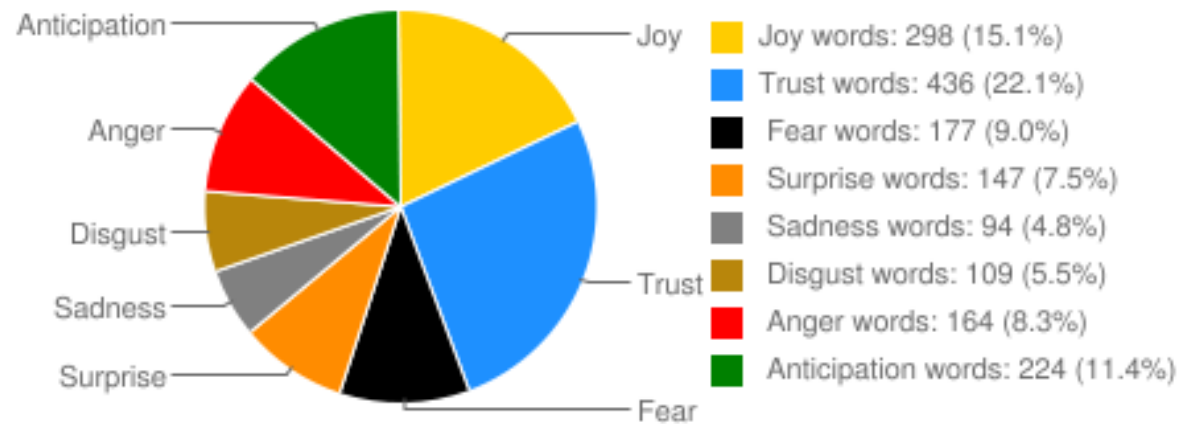
Applications of emotion analysis of books

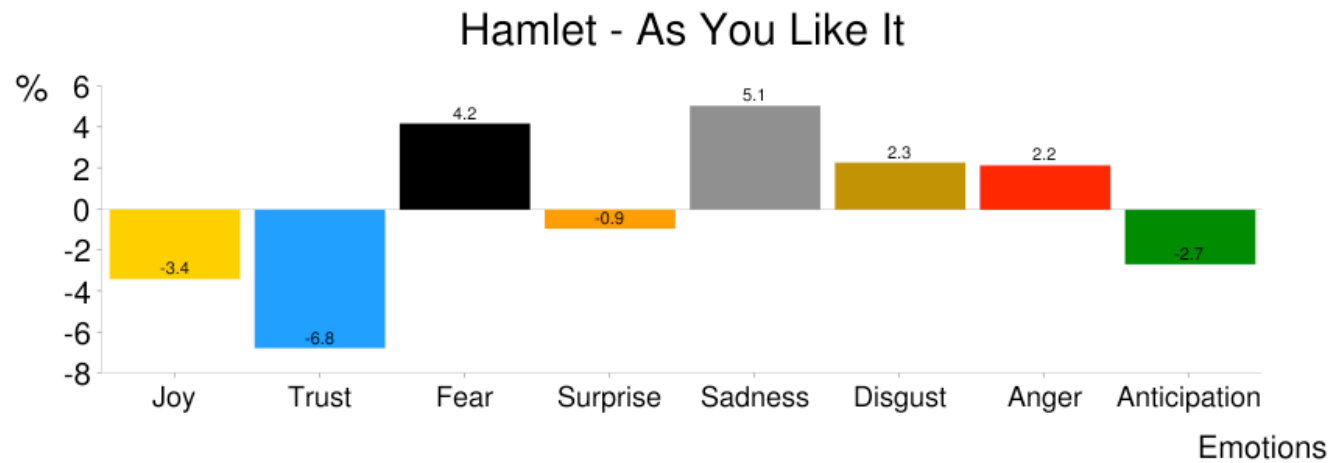
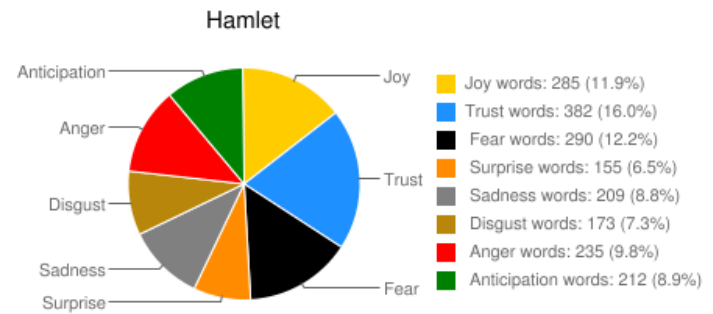
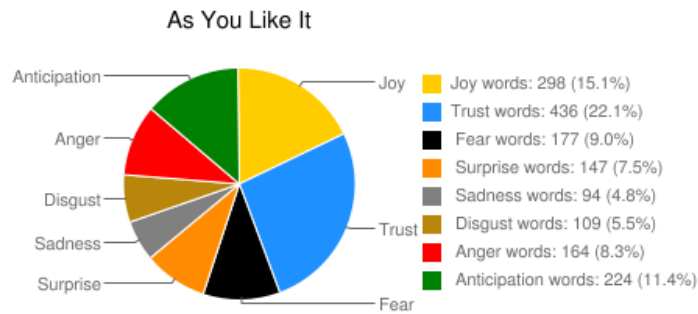
- • Search
 - Example: Which Brothers Grimm tales are the darkest?
- • Social Analysis
 - Example: How have books portrayed entities over time?
(Michel et al. 2011)
- • Literary Analysis
 - Example: Is the distribution of emotion words in fairy tales significantly different from that in novels?
- Summarization
 - Example: Automatically generate summaries that capture different emotional states of characters in a novel
- Analyzing Persuasion Tactics
 - Example: how emotion words are used for persuasion?
(Mannix, 1992; Bales, 1997)

Hamlet

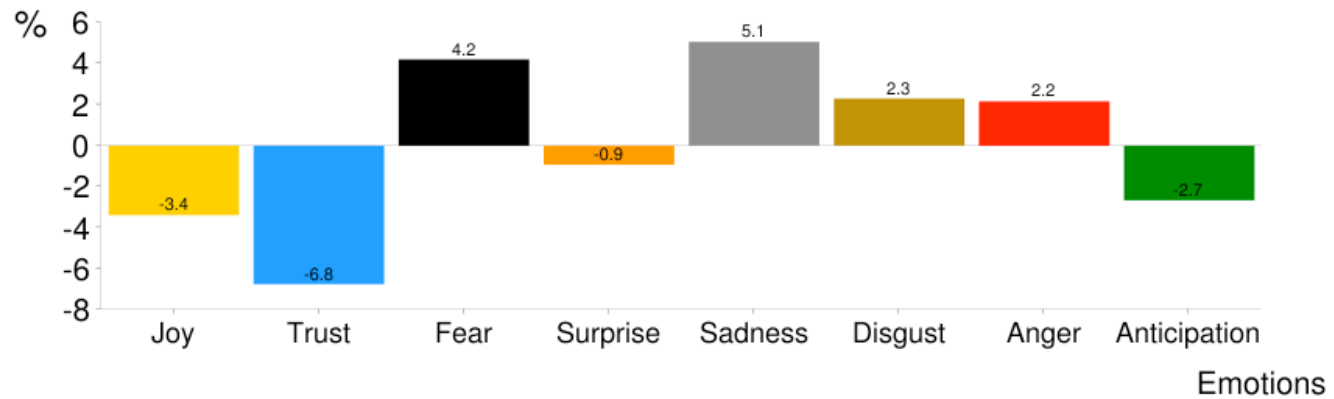


As You Like It





Hamlet - As You Like It



servant esteem Sir **brother** marriage comfort
 loving marry promise fortune virtuous smile
 wonderful oath worthy money hope found remains faithful
 tree honesty friendship **lover** sing synod respect
 proud heavenly praise wear counsel perceive provide
 wealth **pretty** church virgin perfect constant elder invite

relative salience of trust words

soldier sick beating **buried** forfeit doomsday

death malicious guilty confine **grief**

woe sorrow defeated **late** surrender scarcely

suppress **doubt** lose beg black mourning slaughter

frailty mourn **dreadful** **hell** loss shame perilous pious

hideous forbid prison **murder** fat witchcraft

shameful wretch **cursed** disappointed pernicious **mad**

shatter wreck **jealousy** sickness sadness wail sadly

slave confession sterile **tragedy** dismal gore hellish

unequal senseless crash prisoner bleeding wan **drown**

coward oppression drab **devil** affront **affliction** heartache

oppressor **plague** neglected tempest grieve barren suffering

guilt brute forgotten **poison** lament ashamed discomfort debt

murderer weeds dire retirement diseased lowest curse

sickly humble **feeling** nasty **evil** **scourge** disease offender

departed inter damnation bier **rue** wither **burial** ulcer remiss

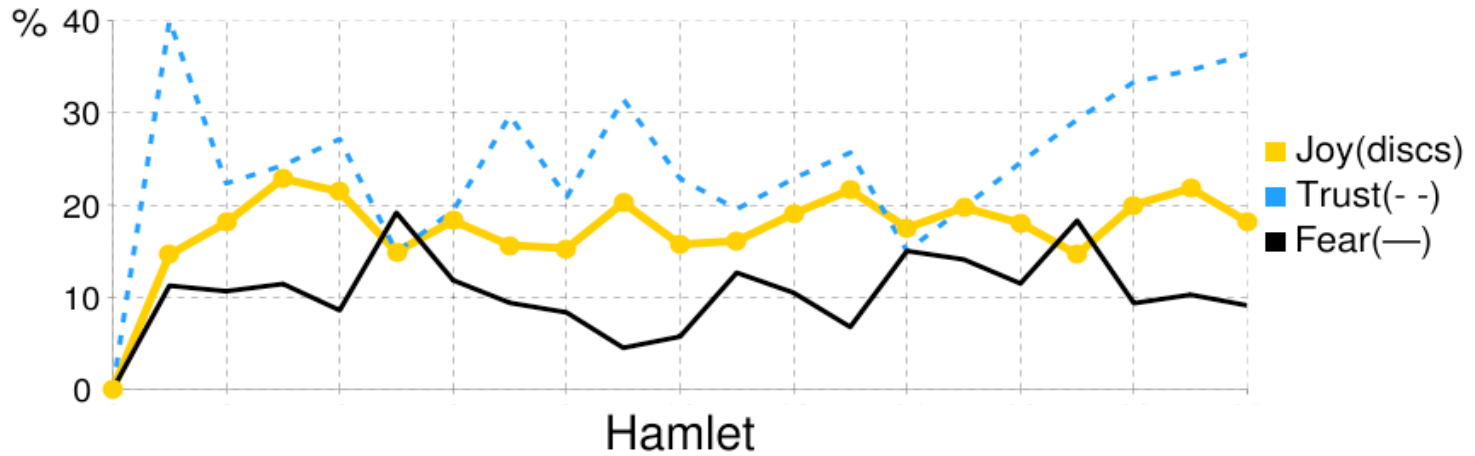
gallows ache losing procession whine perdition shell defy

treachery murderous liquor dying

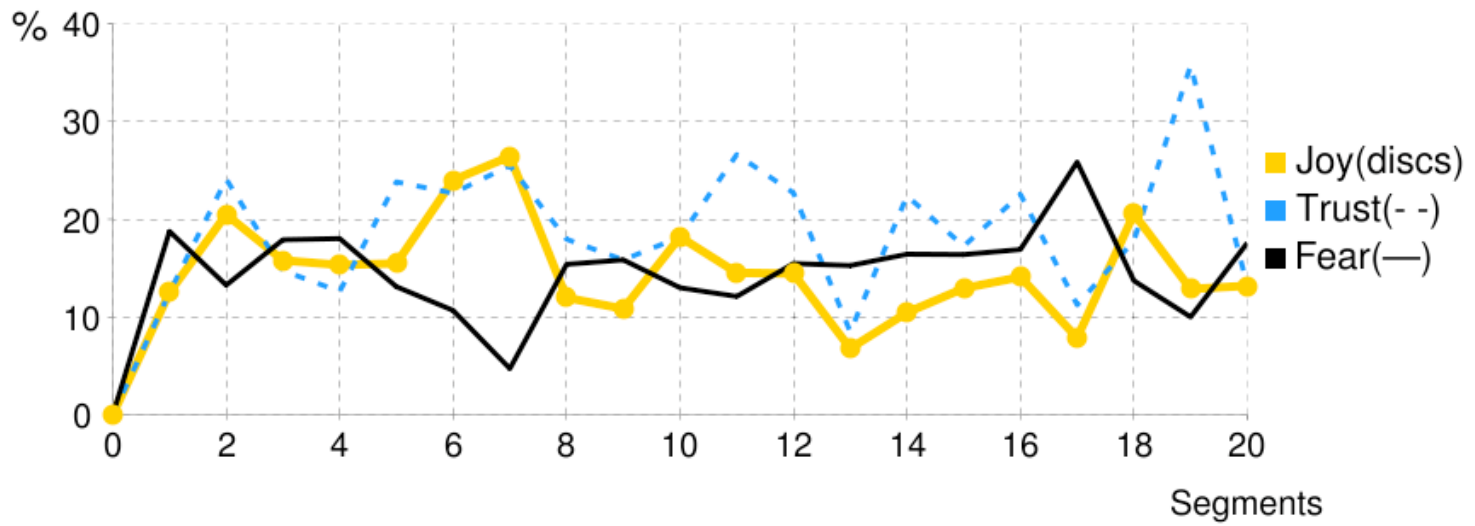
relative salience of sadness words

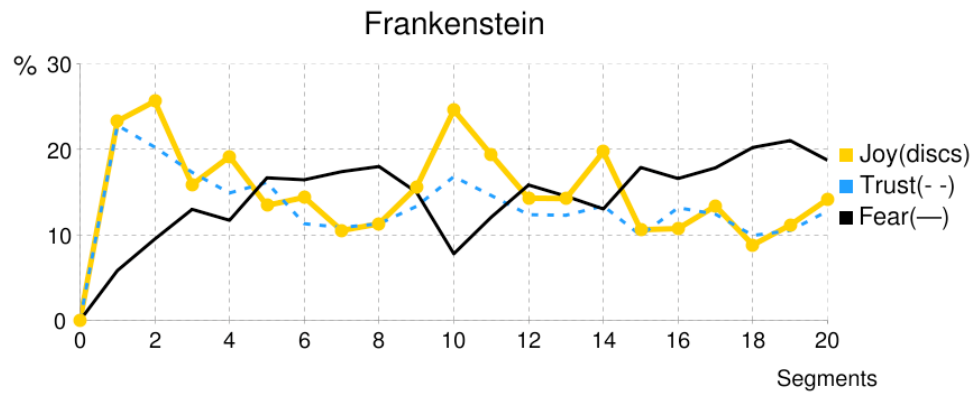
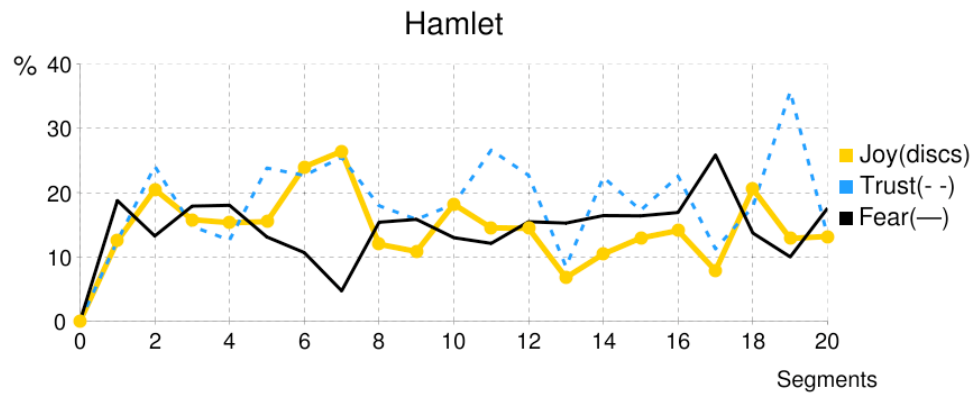
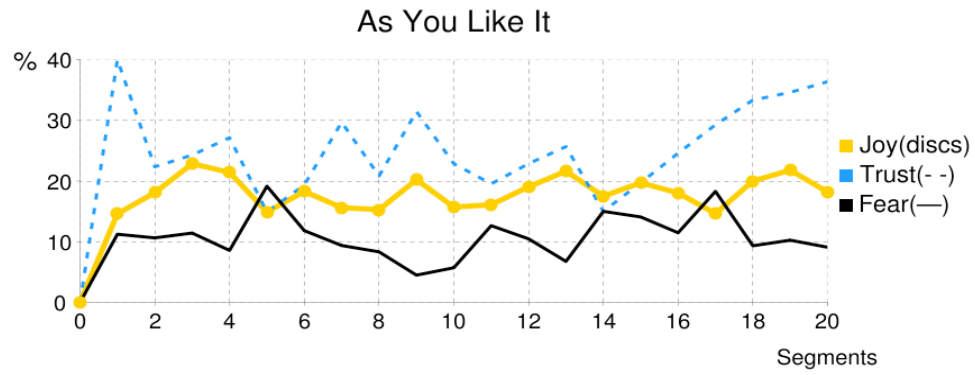
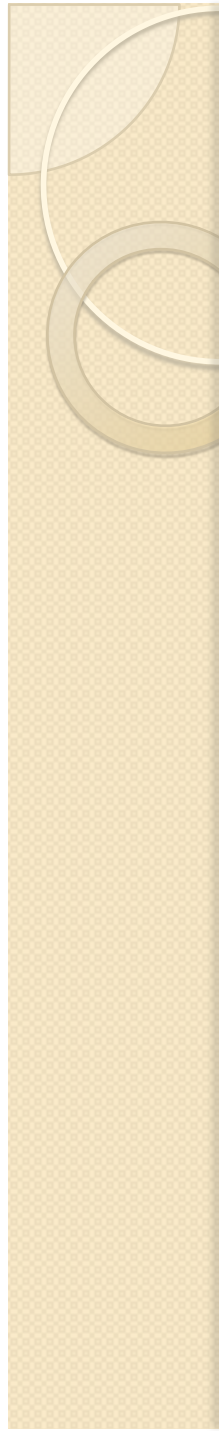
Flow of Emotions

As You Like It



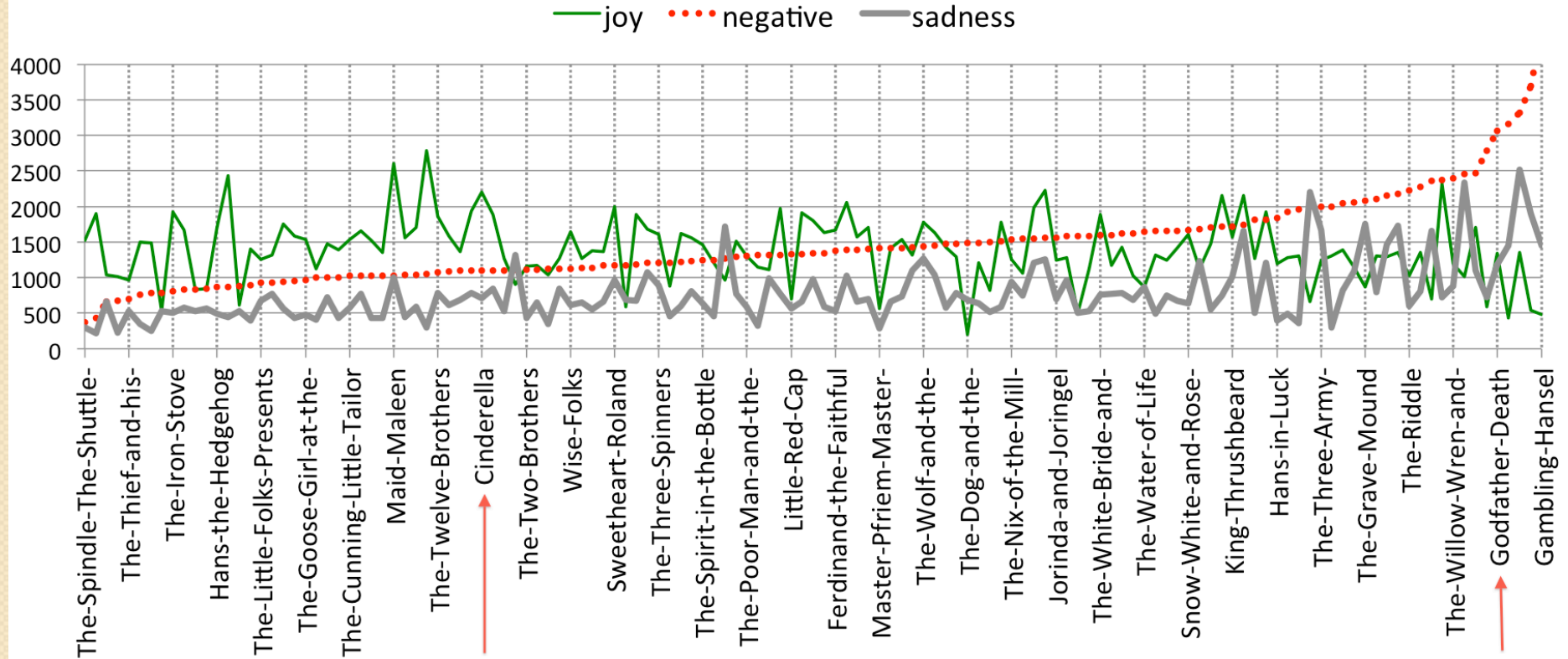
Hamlet





Emotion Word Density

average number of emotion words in every X words

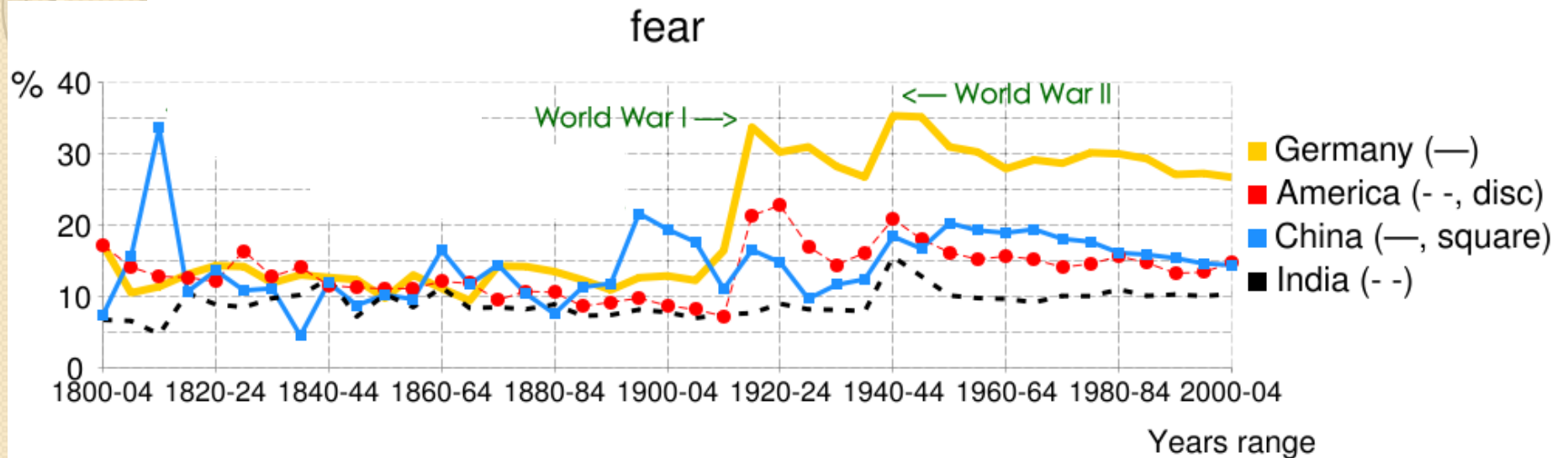


Brothers Grimm fairy tales ordered as per increasing negative word density. X = 10,000.

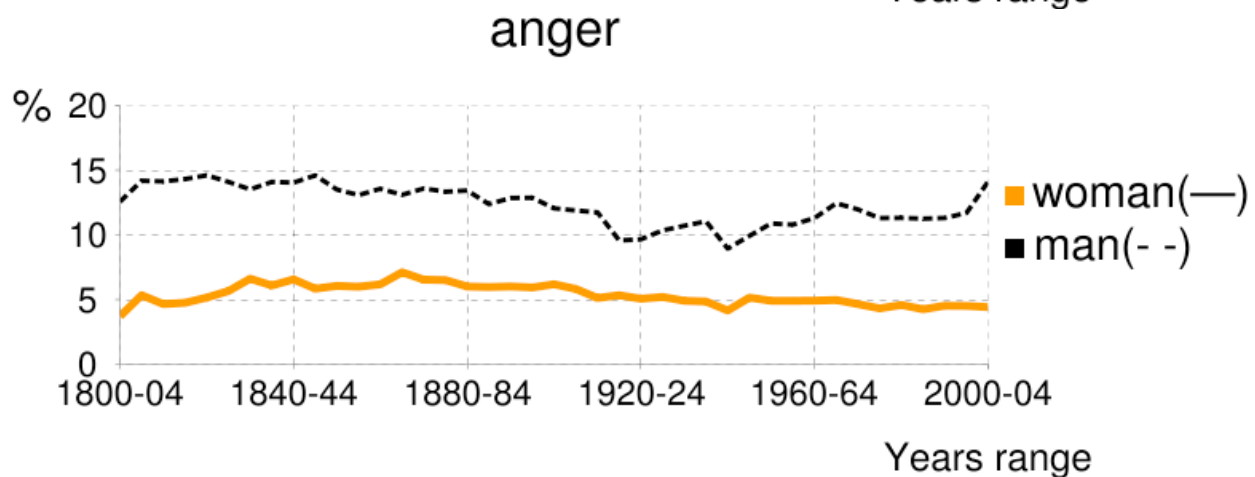
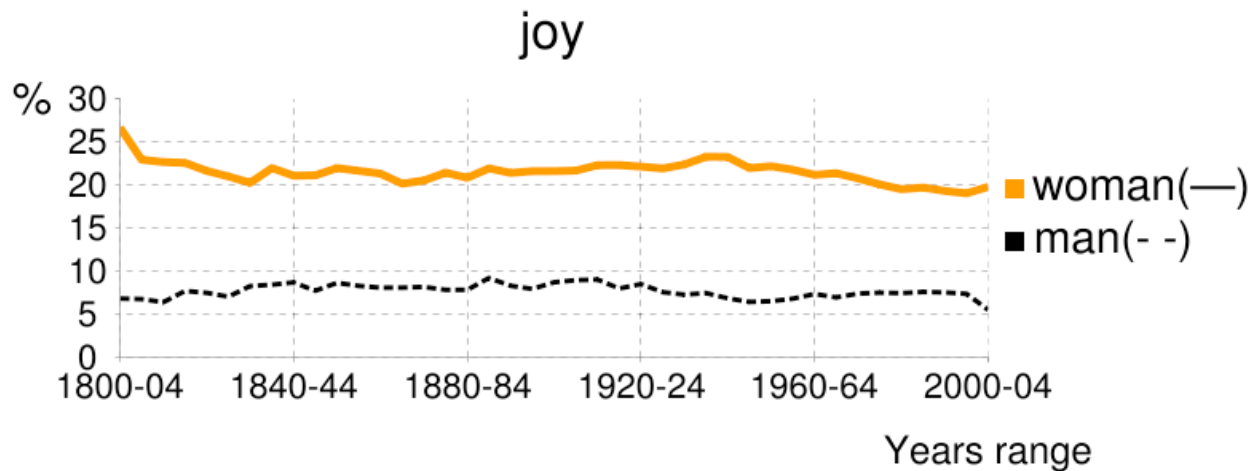


Co-occurring Emotion Words

- Examined emotion words in proximity of target entities
- Used the Google Books Corpus
 - Looked for emotion words in 5-grams that had the target
 - Ignored emotion associated with target word
 - Grouped information into 5-year bins



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



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



Summary

- Created a large word-emotion association lexicon
- Used simple measures and visualizations to quantify and track the use of emotion words in texts
- Used the [Enron Corpus](#)
 - To show use of emotion words by men and women in a work-place setting
- Used the [Brothers Grimm fairy tales](#)
 - showed texts can be ordered for affect-based search
- Used the [Google Books Corpus](#)
 - tracked emotion associations of entities over time