

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Lexicons for Sentiment,  
Affect, and Connotation

# Affective meaning

Drawing on literatures in

- affective computing (Picard 95)
- linguistic subjectivity (Wiebe and colleagues)
- social psychology (Pennebaker and colleagues)

Can we model the lexical semantics relevant to:

- sentiment
- emotion
- personality
- mood
- attitudes

# Why compute affective meaning?

## Detecting:

- sentiment towards politicians, products, countries, ideas
- frustration of callers to a help line
- stress in drivers or pilots
- depression and other medical conditions
- confusion in students talking to e-tutors
- emotions in novels (e.g., for studying groups that are feared over time)

## Could we generate:

- emotions or moods for literacy tutors in the children's storybook domain
- emotions or moods for computer games
- personalities for dialogue systems to match the user

# Connotation in the lexicon

Words have connotation as well as sense

Can we build lexical resources that represent these connotations?

And use them in these computational tasks?



# Scherer Typology of Affective States

**Emotion:** brief organically synchronized ... evaluation of a major event

- *angry, sad, joyful, fearful, ashamed, proud, elated*

**Mood:** diffuse non-caused low-intensity long-duration change in subjective feeling

- *cheerful, gloomy, irritable, listless, depressed, buoyant*

**Interpersonal stances:** affective stance toward another person in a specific interaction

- *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*

**Attitudes:** enduring, affectively colored beliefs, dispositions towards objects or persons

- *liking, loving, hating, valuing, desiring*

**Personality traits:** stable personality dispositions and typical behavior tendencies

- *nervous, anxious, reckless, morose, hostile, jealous*

# Scherer Typology of Affective States

**Emotion:** brief organically synchronized ... evaluation of a major event

- *angry, sad, joyful, fearful, ashamed, proud, elated*

**Mood:** diffuse non-caused low-intensity long-duration change in subjective feeling

- *cheerful, gloomy, irritable, listless, depressed, buoyant*

**Interpersonal stances:** affective stance toward another person in a specific interaction

- *friendly, flirtatious, distant, cold, warm, supportive, contemptuous*

**Attitudes:** enduring, affectively colored beliefs, dispositions towards objects or persons

- *liking, loving, hating, valuing, desiring*

**Personality traits:** stable personality dispositions and typical behavior tendencies

- *nervous, anxious, reckless, morose, hostile, jealous*

# What is a Lexicon?

A (usually hand-built) list of words that correspond to some meaning or class

Possibly with numeric values

Commonly used as simple classifiers, or as features to more complex classifiers

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Lexicons for Sentiment,  
Affect, and Connotation

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Emotion

# Scherer's typology of affective states

**Emotion:** relatively brief episode of synchronized response of all or most organismic subsystems in response to the evaluation of an event as being of major significance

angry, sad, joyful, fearful, ashamed, proud, desperate

**Mood:** diffuse affect state ...change in subjective feeling, of low intensity but relatively long duration, often without apparent cause

cheerful, gloomy, irritable, listless, depressed, buoyant

**Interpersonal stance:** affective stance taken toward another person in a specific interaction, coloring the interpersonal exchange

distant, cold, warm, supportive, contemptuous

**Attitudes:** relatively enduring, affectively colored beliefs, preferences predispositions towards objects or persons

liking, loving, hating, valuing, desiring

**Personality traits:** emotionally laden, stable personality dispositions and behavior tendencies, typical for a person

nervous, anxious, reckless, morose, hostile, envious, jealous

# Two families of theories of emotion

## Atomic basic emotions

- A finite list of 6 or 8, from which others are generated

## Dimensions of emotion

- Valence (positive negative)
- Arousal (strong, weak)
- Control



# Ekman's 6 basic emotions:

## Surprise, happiness, anger, fear, disgust, sadness

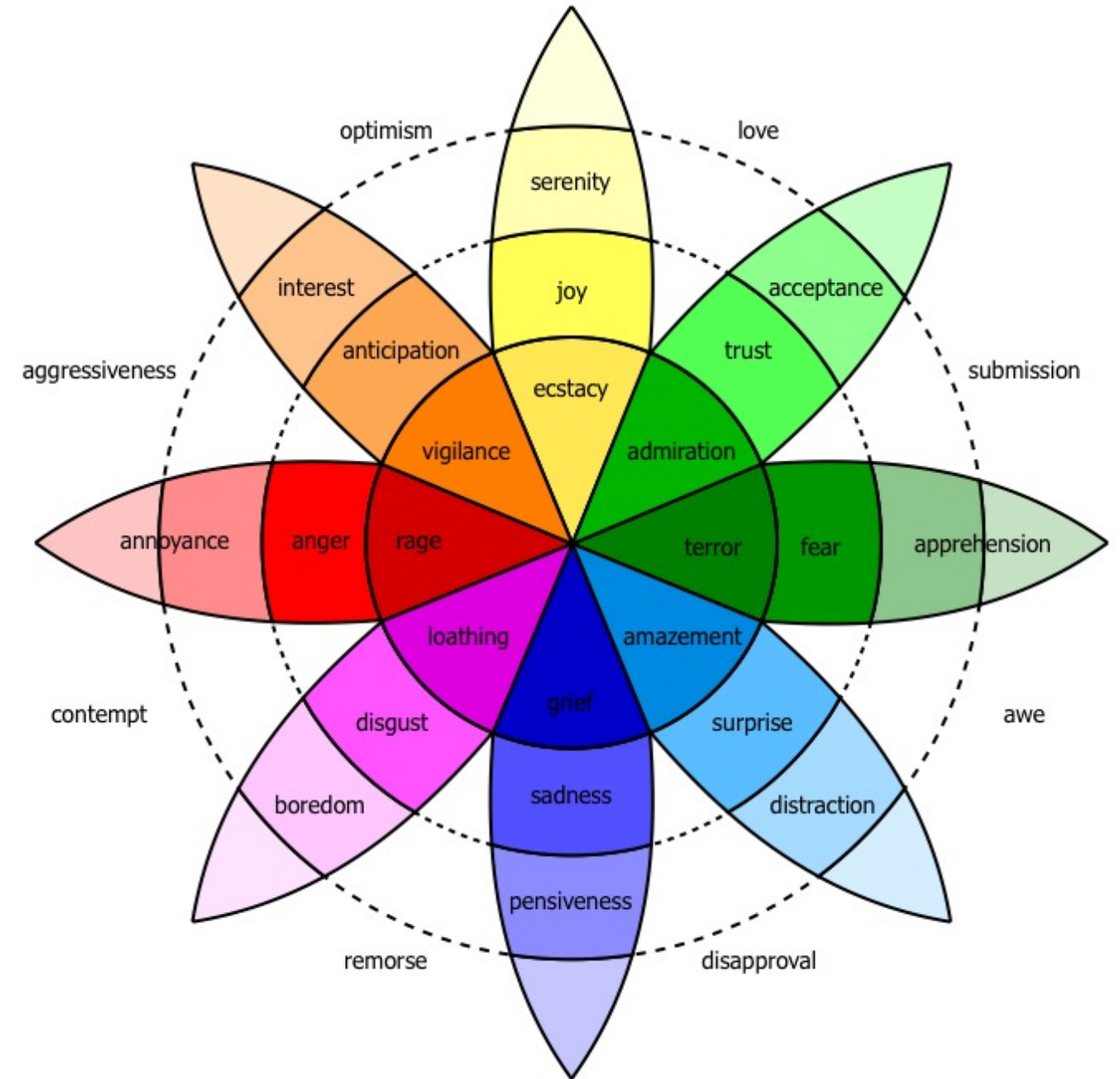


Ekman &  
Matsumoto  
1989



# Plutchick's wheel of em

- 8 basic emotions
- in four opposing pairs:
  - joy–sadness
  - anger–fear
  - trust–disgust
  - anticipation–surprise



# Alternative: spatial model

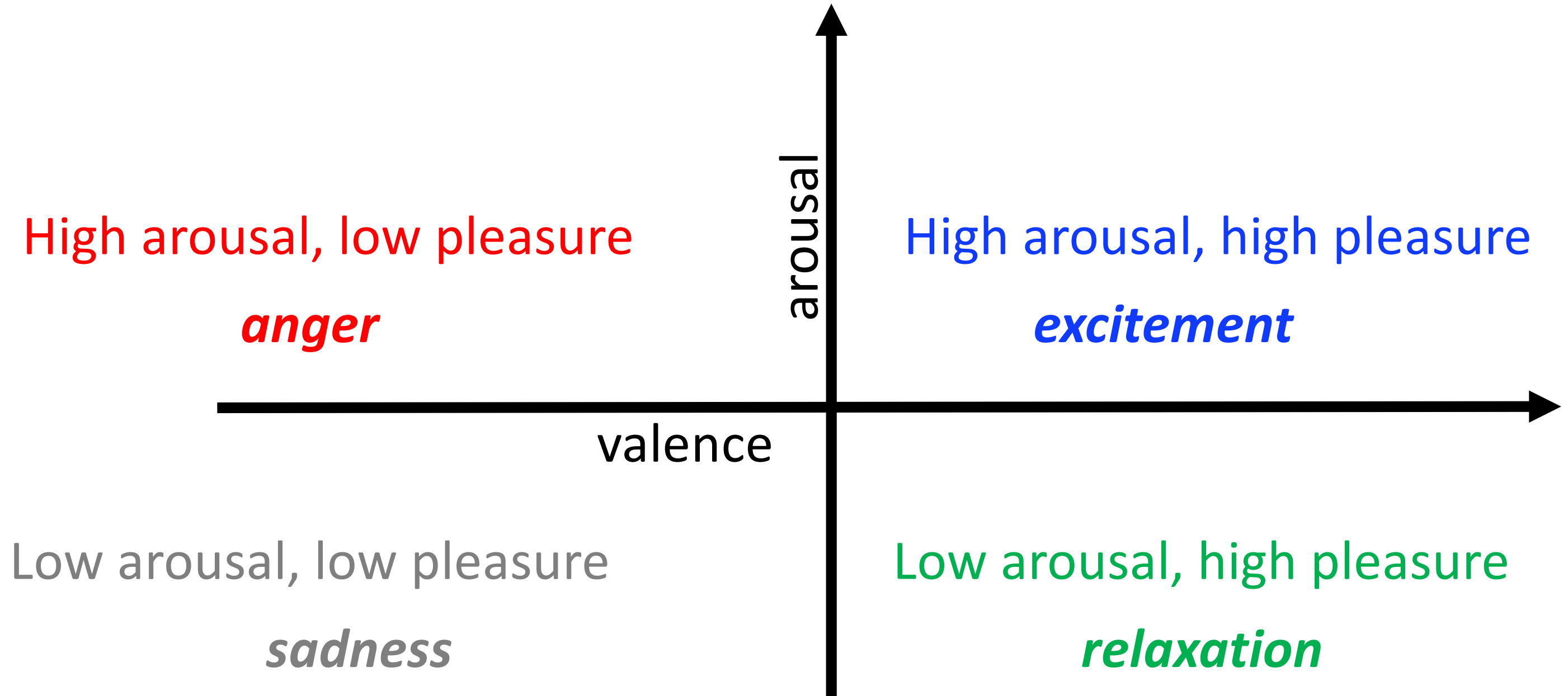
An emotion is a point in 2- or 3-dimensional space

**valence:** the pleasantness of the stimulus

**arousal:** the intensity of emotion provoked by the stimulus

(sometimes) **dominance:** the degree of control exerted by the stimulus

# Valence/Arousal Dimensions



# Simple sentiment lexicons

# The General Inquirer

Philip J. Stone, Dexter C Dunphy, Marshall S. Smith,  
Daniel M. Ogilvie. 1966. The General Inquirer: A  
Computer Approach to Content Analysis. MIT Press

Positiv (1915 words)  
Negativ (2291 words)

| 1    | Entry     | Source | Positiv | Negativ |
|------|-----------|--------|---------|---------|
| 2586 | DAKOTA    | Lvd    |         |         |
| 2587 | DAMAGE#1  | H4Lvd  |         | Negativ |
| 2588 | DAMAGE#2  | H4Lvd  |         | Negativ |
| 2589 | DAMN      | H4Lvd  |         | Negativ |
| 2590 | DAMNABLE  | H4     |         | Negativ |
| 2591 | DAMNED    | H4     |         | Negativ |
| 2592 | DAMP      | H4Lvd  |         |         |
| 2593 | DANCE#1   | H4Lvd  | Positiv |         |
| 2594 | DANCE#2   | H4Lvd  | Positiv |         |
| 2595 | DANCE#3   | H4Lvd  | Positiv |         |
| 2596 | DANCER    | H4Lvd  |         |         |
| 2597 | DANGER    | H4Lvd  |         | Negativ |
| 2598 | DANGEROUS | H4Lvd  |         | Negativ |
| 2599 | DANISH    | Lvd    |         |         |
| 2600 | DARE      | H4Lvd  | Positiv |         |
| 2601 | DARING    | H4Lvd  | Positiv |         |
| 2602 | DARK      | H4Lvd  |         | Negativ |
| 2603 | DARKEN    | H4Lvd  |         | Negativ |
| 2604 | DARKNESS  | H4Lvd  |         | Negativ |
| 2605 | DARLING   | H4Lvd  | Positiv |         |

# MPQA Subjectivity Cues Lexicon

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. Proc. of HLT-EMNLP-2005.

Riloff and Wiebe (2003). Learning extraction patterns for subjective expressions. EMNLP-2003.

6885 words

Is a subjective word positive or negative?

- Strongly or weakly?

<http://mpqa.cs.pitt.edu/lexicons/>

GNU GPL

type=**weaksubj** len=1 word1=**abandoned** pos1=adj stemmed1=n priorpolarity=**negative**  
type=weaksubj len=1 word1=**abandonment** pos1=noun stemmed1=n priorpolarity=negative  
type=weaksubj len=1 word1=**abandon** pos1=verb stemmed1=y priorpolarity=negative  
type=strongsubj len=1 word1=**abase** pos1=verb stemmed1=y priorpolarity=negative  
type=**strongsubj** len=1 word1=**abasement** pos1=anypos stemmed1=y priorpolarity=negative  
type=strongsubj len=1 word1=**abash** pos1=verb stemmed1=y priorpolarity=negative  
type=weaksubj len=1 word1=**abate** pos1=verb stemmed1=y priorpolarity=negative



# Words with consistent sentiment across lexicons

|                 |   |
|-----------------|---|
| <b>Positive</b> | admire, amazing, assure, celebration, charm, eager, enthusiastic, excellent, fancy, fantastic, frolic, graceful, happy, joy, luck, majesty, mercy, nice, patience, perfect, proud, rejoice, relief, respect, satisfactorily, sensational, super, terrific, thank, vivid, wise, wonderful, zest        |
| <b>Negative</b> | abominable, anger, anxious, bad, catastrophe, cheap, complaint, condescending, deceit, defective, disappointment, embarrass, fake, fear, filthy, fool, guilt, hate, idiot, inflict, lazy, miserable, mourn, nervous, objection, pest, plot, reject, scream, silly, terrible, unfriendly, vile, wicked |



# Let's look at two emotion lexicons!

## 1. 8 basic emotions:

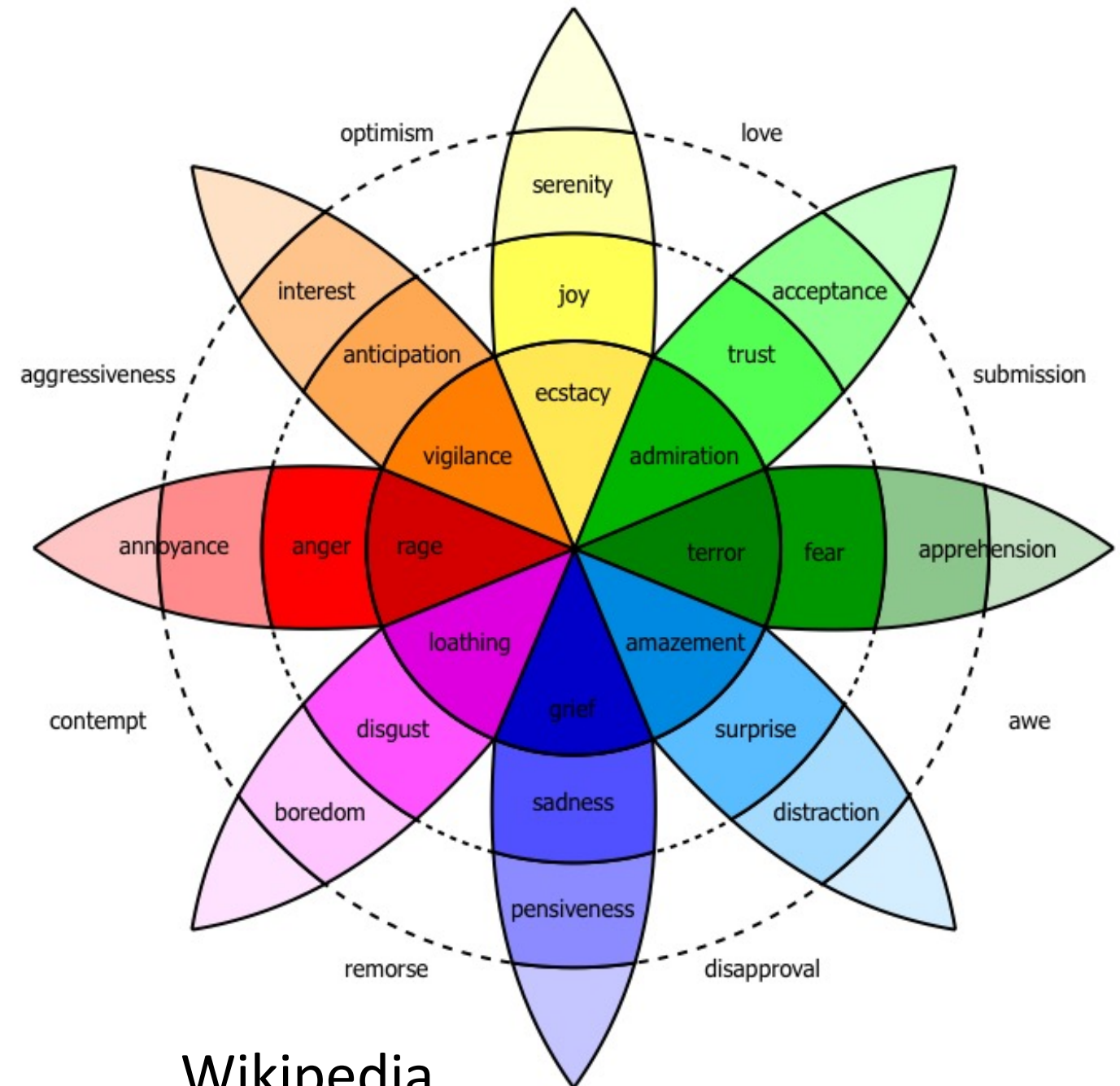
- NRC Word-Emotion Association Lexicon (Mohammad and Turney 2011)

## 2. Dimensions of valence/arousal/dominance

- NRC Valence-Arousal-Dominance Lexicon (Mohammad 2018)

# Plutchick's wheel of emotion

- 8 basic emotions
- in four opposing pairs:
  - joy–sadness
  - anger–fear
  - trust–disgust
  - anticipation–surprise



Wikipedia

# NRC Word-Emotion Association Lexicon

Mohammad and Turney 2011

|           |              |   |   |
|-----------|--------------|---|---|
| amazingly | anger        | 0 |   |
| amazingly | anticipation | 0 |   |
| amazingly | disgust      | 0 |   |
| amazingly | fear         | 0 |   |
| amazingly | joy          | 1 |   |
| amazingly | sadness      | 0 |   |
| amazingly | surprise     |   | 1 |
| amazingly | trust        | 0 |   |
| amazingly | negative     |   | 0 |
| amazingly | positive     |   | 1 |

# More examples

| Word       | anger | anticipation | disgust | fear | joy | sadness | surprise | trust | positive | negative |
|------------|-------|--------------|---------|------|-----|---------|----------|-------|----------|----------|
| reward     | 0     | 1            | 0       | 0    | 1   | 0       | 1        | 1     | 1        | 0        |
| worry      | 0     | 1            | 0       | 1    | 0   | 1       | 0        | 0     | 0        | 1        |
| tenderness | 0     | 0            | 0       | 0    | 1   | 0       | 0        | 0     | 1        | 0        |
| sweetheart | 0     | 1            | 0       | 0    | 1   | 1       | 0        | 1     | 1        | 0        |
| suddenly   | 0     | 0            | 0       | 0    | 0   | 0       | 1        | 0     | 0        | 0        |
| thirst     | 0     | 1            | 0       | 0    | 0   | 1       | 1        | 0     | 0        | 0        |
| garbage    | 0     | 0            | 1       | 0    | 0   | 0       | 0        | 0     | 0        | 1        |

# NRC Emotion/Affect Intensity Lexicon (Mohammad, 2018b); real values for 5814 words

| Anger      |       | Fear       |       | Joy      |       | Sadness      |       |
|------------|-------|------------|-------|----------|-------|--------------|-------|
| outraged   | 0.964 | horror     | 0.923 | superb   | 0.864 | sad          | 0.844 |
| violence   | 0.742 | anguish    | 0.703 | cheered  | 0.773 | guilt        | 0.750 |
| coup       | 0.578 | pestilence | 0.625 | rainbow  | 0.531 | unkind       | 0.547 |
| oust       | 0.484 | stressed   | 0.531 | gesture  | 0.387 | difficulties | 0.421 |
| suspicious | 0.484 | failing    | 0.531 | warms    | 0.391 | beggar       | 0.422 |
| nurture    | 0.059 | confident  | 0.094 | hardship | .031  | sing         | 0.017 |

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Other Useful Lexicons

# LIWC: Linguistic Inquiry and Word Count

| Positive Emotion | Negative Emotion | Insight  | Inhibition | Family    | Negate  |
|------------------|------------------|----------|------------|-----------|---------|
| appreciat*       | anger*           | aware*   | avoid*     | brother*  | aren't  |
| comfort*         | bore*            | believe  | careful*   | cousin*   | cannot  |
| great            | cry              | decid*   | hesitat*   | daughter* | didn't  |
| happy            | despair*         | feel     | limit*     | family    | neither |
| interest         | fail*            | figur*   | oppos*     | father*   | never   |
| joy*             | fear             | know     | prevent*   | grandf*   | no      |
| perfect*         | griev*           | knew     | reluctan*  | grandm*   | nobod*  |
| please*          | hate*            | means    | safe*      | husband   | none    |
| safe*            | panic*           | notice*  | stop       | mom       | nor     |
| terrific         | suffers          | recogni* | stubborn*  | mother    | nothing |
| value            | terrify          | sense    | wait       | niece*    | nowhere |
| wow*             | violent*         | think    | wary       | wife      | without |

# LIWC (Linguistic Inquiry and Word Count)

Pennebaker, J.W., Booth, R.J., & Francis, M.E. (2007). Linguistic Inquiry and Word Count: LIWC 2007. Austin, TX

<http://www.liwc.net/>

2300 words

>70 classes



# The General Inquirer

Philip J. Stone, Dexter C Dunphy, Marshall S. Smith, Daniel M. Ogilvie. 1966. The General Inquirer: A Computer Approach to Content Analysis. MIT Press

- Home page: <http://www.wjh.harvard.edu/~inquirer>
- List of Categories: <http://www.wjh.harvard.edu/~inquirer/homecat.htm>
- Spreadsheet: <http://www.wjh.harvard.edu/~inquirer/inquirerbasic.xls>

## Categories:

- Positiv (1915 words) and Negativ (2291 words)
- Strong vs Weak, Active vs Passive, Overstated versus Understated
- **Pleasure, Pain, Virtue, Vice, Motivation, Cognitive Orientation, etc**

Free for Research Use

# Concreteness versus abstractness

The degree to which the concept denoted by a word refers to a perceptible entity.

Brysbaert, M., Warriner, A. B., and Kuperman, V. (2014) [Concreteness ratings for 40 thousand generally known English word lemmas](#) *Behavior Research Methods* 46, 904-911.

[Supplementary data: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.](#)

37,058 English words and 2,896 two-word expressions ( “zebra crossing” and “zoom in”),

Rating from 1 (abstract) to 5 (concrete)

# Concreteness versus abstractness

Brysbaert, M., Warriner, A. B., and Kuperman, V. (2014) [Concreteness ratings for 40 thousand generally known English word lemmas \*Behavior Research Methods\* 46, 904-911.](#)

[Supplementary data: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License.](#)

Some example ratings from the final dataset of 40,000 words and phrases

banana 5

bathrobe 5

bagel 5

brisk 2.5

badass 2.5

basically 1.32

belief 1.19

although 1.07

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Building Lexicons using  
Human Labelers

# Where do lexicons come from?

- One method: crowdsourcing!!!
- 10,000 words
  - Collected from earlier lexicons
- Labeled by workers on Amazon Mechanical Turk
  - “Turkers”
- 5 Turkers per hit

# The AMT Hit

Q4. How much is *startle* associated with the emotion joy? (For example, *happy* and *fun* are strongly associated with joy.)

- *startle* is not associated with joy
- *startle* is weakly associated with joy
- *startle* is moderately associated with joy
- *startle* is strongly associated with joy

Q5. How much is *startle* associated with the emotion sadness? (For example, *failure* and *heart-break* are strongly associated with sadness.)

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

Q6. How much is *startle* associated with the emotion fear? (For example, *horror* and *scary* are strongly associated with fear.)

...

# NRC Valence, Arousal, Dominance (VAD) lexicon

Mohammad (2018)

20,000 words, 3 emotional dimensions:

- **valence** (the pleasantness of the stimulus)
- **arousal** (the intensity of emotion provoked by the stimulus)
- **dominance** (the degree of control exerted by the stimulus)

# Best-worst scaling: valence

Q1. Which of the four words below is associated with the MOST happiness / pleasure / positiveness / satisfaction / contentedness / hopefulness OR LEAST unhappiness / annoyance / negativeness / dissatisfaction / melancholy / despair?

vacation, consolation, whistle, torture

Q2. Which of the four words below is associated with the LEAST happiness / pleasure / positiveness / satisfaction / contentedness / hopefulness OR MOST unhappiness / annoyance / negativeness / dissatisfaction / melancholy / despair?



# Lexicon of valence, arousal, and dominance

| Valence     |      | Arousal    |      | Dominance   |       |
|-------------|------|------------|------|-------------|-------|
| delightful  | .918 | enraged    | .962 | powerful    | .991  |
| vacation    | .840 | party      | .840 | authority   | .935  |
| whistle     | .653 | organized  | .337 | saxophone   | .482  |
| consolation | .408 | effortless | .120 | discouraged | .0090 |
| torture     | .115 | napping    | .046 | weak        | .045  |

# Issues to keep in mind with crowdsourcing lexicons

Native (or very fluent) speakers

Making the task clear for non-linguists or non computer scientists

Paying minimum wage ([fairwork.stanford.edu](http://fairwork.stanford.edu))

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Building Lexicons using  
Human Labelers

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Semi-supervised Induction  
of Affect Lexicons

# Semantic Axis Methods

(An et al., 2018, Turney and Littman 2003)

Start with seed words like *good* or *bad* for the two poles

For each word to be added to lexicon

- Compute a word representation
- Use this to measure its distance from the poles
- Assign it to the pole it is closer to

# Initial seeds for different domains

- (1) Start with a single large seed lexicon and rely on the induction algorithm to fine-tune it to the domain
- (2) Choose different seed words for different genres:

| Domain  | Positive seeds   | Negative seeds  |
|---------|--|---|
| General | good, lovely, excellent, fortunate, pleasant, delightful, perfect, loved, love, happy    | bad, horrible, poor, unfortunate, unpleasant, disgusting, evil, hated, hate, unhappy        |
| Twitter | love, loved, loves, awesome, nice, amazing, best, fantastic, correct, happy              | hate, hated, hates, terrible, nasty, awful, worst, horrible, wrong, sad                     |
| Finance | successful, excellent, profit, beneficial, improving, improved, success, gains, positive | negligent, loss, volatile, wrong, losses, damages, bad, litigation, failure, down, negative |

# Compute representation

Can just use off-the-shelf static embeddings

- word2vec, GloVe, etc.

Or compute on a corpus

Or fine-tune pre-trained embeddings to a corpus

# Represent each pole

Start with embeddings for seed words:

$$S^+ = \{E(w_1^+), E(w_2^+), \dots, E(w_n^+)\}$$

$$S^- = \{E(w_1^-), E(w_2^-), \dots, E(w_m^-)\}$$

Pole centroids are:

$$\mathbf{V}^+ = \frac{1}{n} \sum_1^n E(w_i^+)$$

$$\mathbf{V}^- = \frac{1}{m} \sum_1^m E(w_i^-)$$

Semantic axis is:

$$\mathbf{V}_{axis} = \mathbf{V}^+ - \mathbf{V}^-$$

Word score is cosine with axis

$$\begin{aligned} \text{score}(w) &= (\cos(E(w), \mathbf{V}_{axis})) \\ &= \frac{E(w) \cdot \mathbf{V}_{axis}}{\|E(w)\| \|\mathbf{V}_{axis}\|} \end{aligned}$$



# Label Propagation Methods

Alternative to axis methods: propagate sentiment labels on word graphs

First proposed by Hatzivassiloglou and McKeown (1997)

Let's see method of Hamilton et al. 2016

# Label Propagation (Hamilton et al., 2016 version)

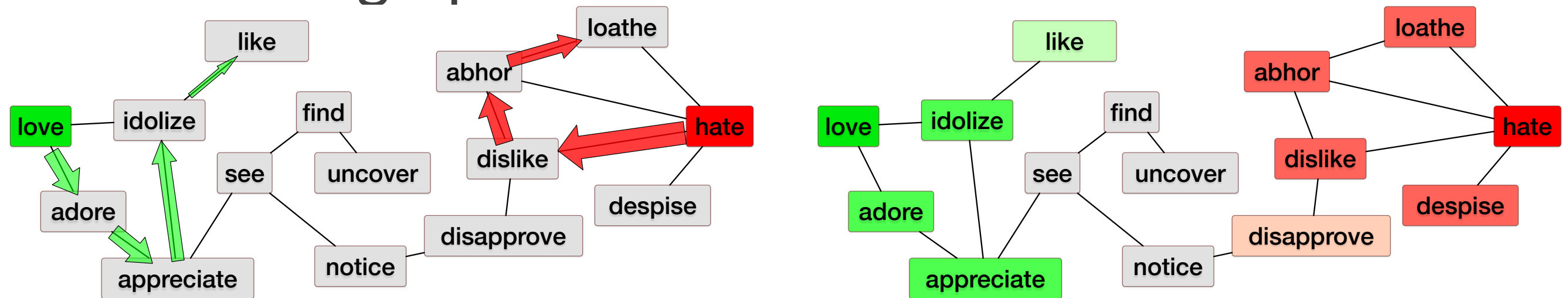
1. Define a graph: connecting each word with  $k$  nearest neighbor

$$\mathbf{E}_{i,j} = \arccos \left( -\frac{\mathbf{w}_i^\top \mathbf{w}_j}{\|\mathbf{w}_i\| \|\mathbf{w}_j\|} \right)$$

2. Define a seed set (pos and neg words)  
love, hate, etc

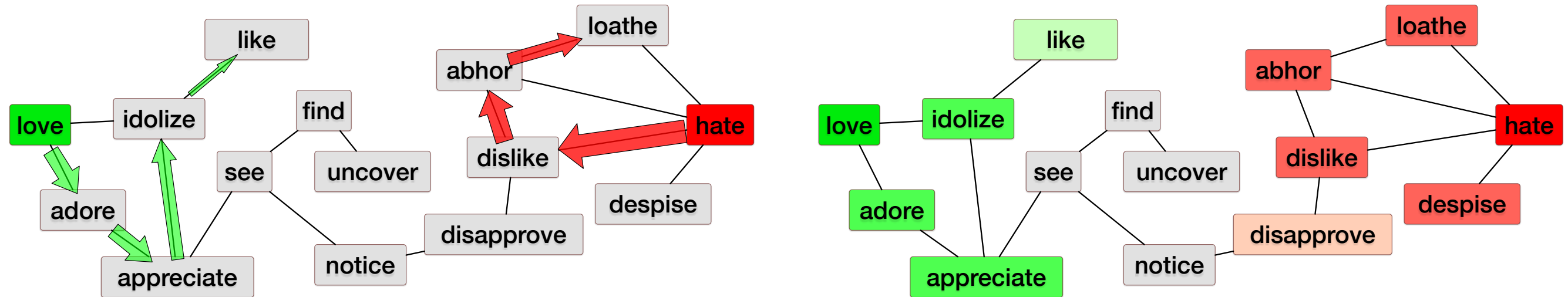
# Label Propagation (Hamilton et al., 2016 version)

## 3. Propagate polarities from the seed set: randomly walk on the graph



Polarity score is proportional to probability of random walk landing on word

# Label Propagation (Hamilton et al., 2016 version)

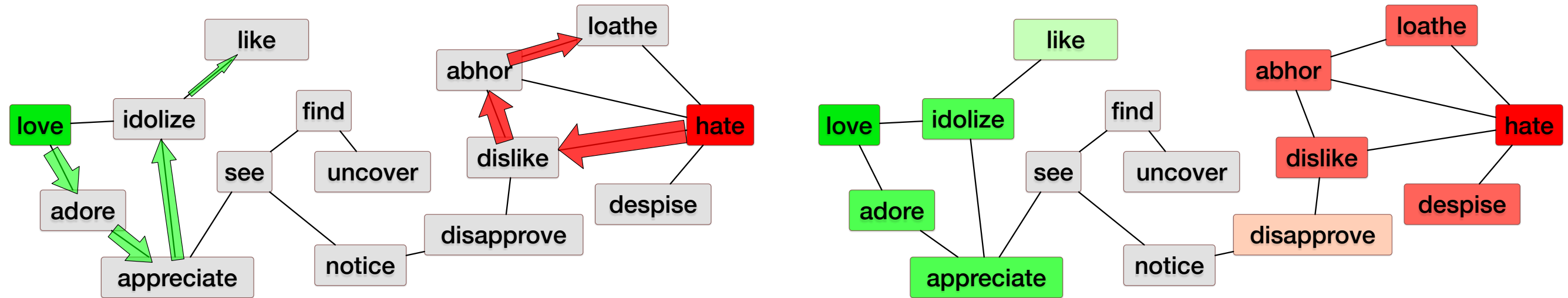


## 4. Create word scores:

- Walking from positive and negative seedsets
  - Gives  $\text{rawscore}^+(w_i)$  and  $\text{rawscore}^-(w_i)$
- Combine into one score:

$$\text{score}^+(w_i) = \frac{\text{rawscore}^+(w_i)}{\text{rawscore}^+(w_i) + \text{rawscore}^-(w_i)}$$

# Label Propagation (Hamilton et al., 2016 version)



## 5. Assign confidence via bootstrap sampling:

- Compute the propagation  $B$  times over random subsets of the positive and negative seed sets
- The standard deviation of the bootstrap sampled polarity scores gives a confidence measure.

# Other metrics besides cosine:

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997.  
Predicting the Semantic Orientation of Adjectives. ACL, 174–181

Adjectives conjoined by “*and*” have same polarity

- Fair **and** legitimate, corrupt **and** brutal
- \*fair **and** brutal, \*corrupt **and** legitimate

Adjectives conjoined by “*but*” do not

- fair **but** brutal

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Semi-supervised Induction  
of Affect Lexicons

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Supervised Learning of  
Word Sentiment



# Learn word sentiment supervised by online review scores

Potts, Christopher. 2011. On the negativity of negation. SALT 20, 636-659.  
Potts 2011 NSF Workshop talk.

## Review datasets

- IMDB, Goodreads, Open Table, Amazon, Trip Advisor

Each review has a score (1-5, 1-10, etc)

Just count how many times each word occurs with each score

- (and normalize)

# Online review data

## **Movie review excerpts (IMDb)**

- 10** A great movie. This film is just a wonderful experience. It's surreal, zany, witty and slapstick all at the same time. And terrific performances too.
- 1** This was probably the worst movie I have ever seen. The story went nowhere even though they could have done some interesting stuff with it.

## **Restaurant review excerpts (Yelp)**

- 5** The service was impeccable. The food was cooked and seasoned perfectly... The watermelon was perfectly square ... The grilled octopus was ... mouthwatering...
- 2** ...it took a while to get our waters, we got our entree before our starter, and we never received silverware or napkins until we requested them...

## **Book review excerpts (GoodReads)**

- 1** I am going to try and stop being deceived by eye-catching titles. I so wanted to like this book and was so disappointed by it.
- 5** This book is hilarious. I would recommend it to anyone looking for a satirical read with a romantic twist and a narrator that keeps butting in

## **Product review excerpts (Amazon)**

- 5** The lid on this blender though is probably what I like the best about it... enables you to pour into something without even taking the lid off! ... the perfect pitcher! ... works fantastic.
- 1** I hate this blender... It is nearly impossible to get frozen fruit and ice to turn into a smoothie... You have to add a TON of liquid. I also wish it had a spout ...

# Analyzing the polarity of each word in IMDB

Potts, Christopher. 2011. On the negativity of negation. SALT 20, 636-659.

How likely is each word to appear in each sentiment class?

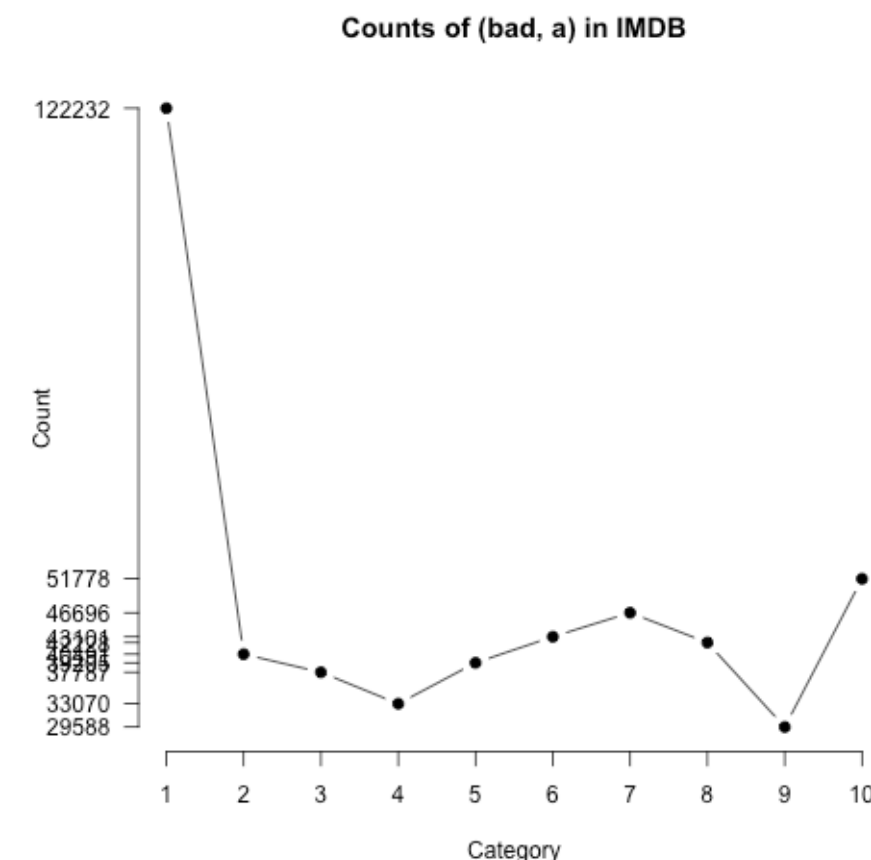
Count("bad") in 1-star, 2-star, 3-star, etc.

But can't use raw counts:

Instead, **likelihood**:  $P(w | c) = \frac{f(w, c)}{\sum_{w \in c} f(w, c)}$

Make them comparable between words

- **Scaled likelihood**:  $\frac{P(w | c)}{P(w)}$

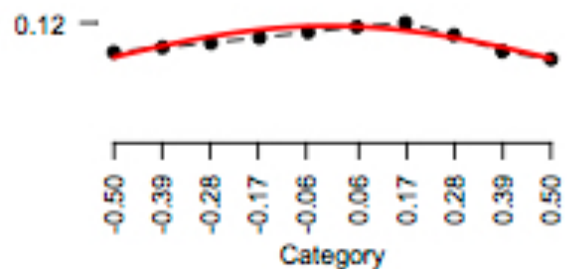


# “Potts diagrams”

Potts, Christopher. 2011. NSF workshop on restructuring adjectives.

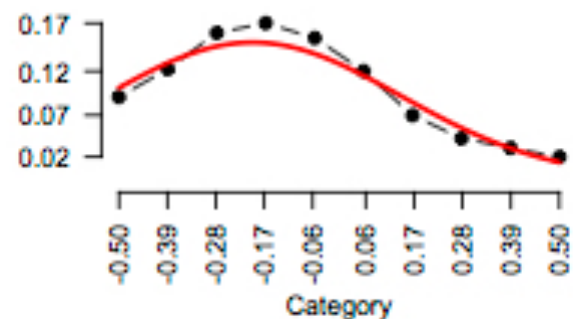
## Positive scalars

good



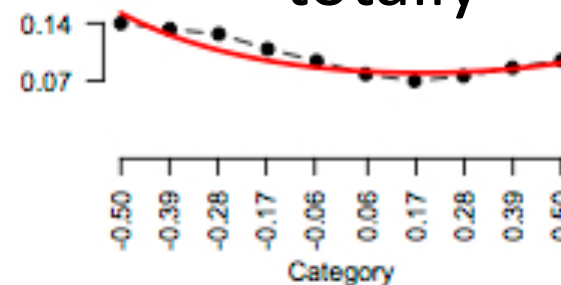
## Negative scalars

disappointing



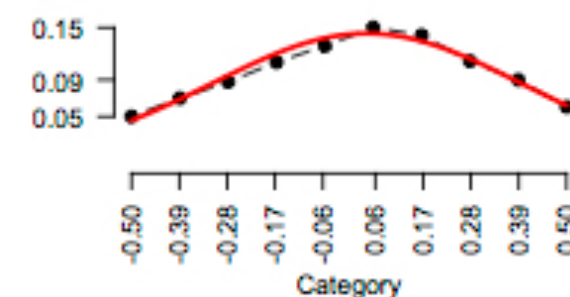
## Emphatics

totally

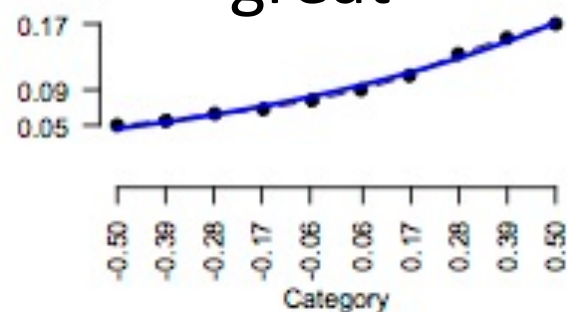


## Attenuators

somewhat



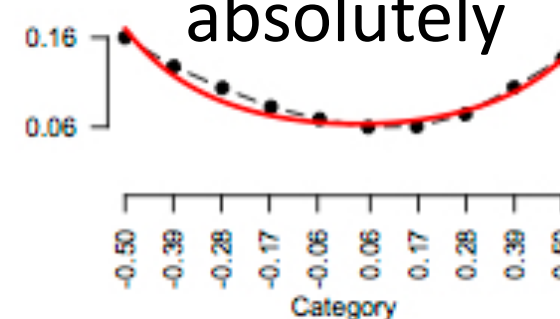
great



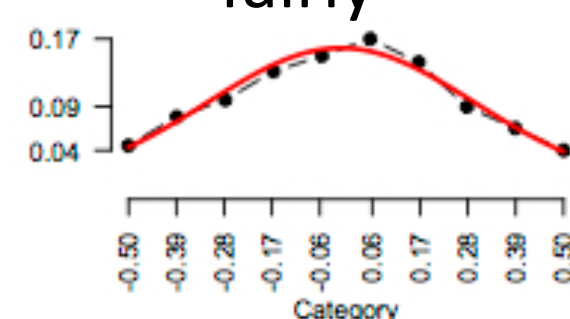
bad



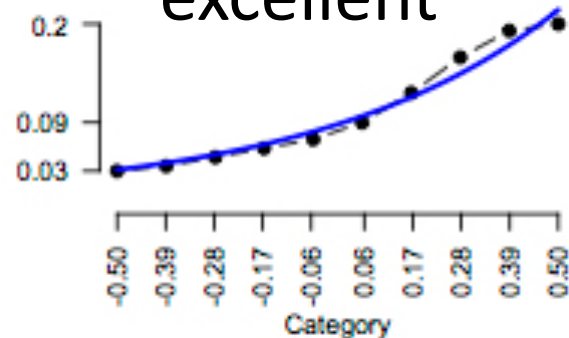
absolutely



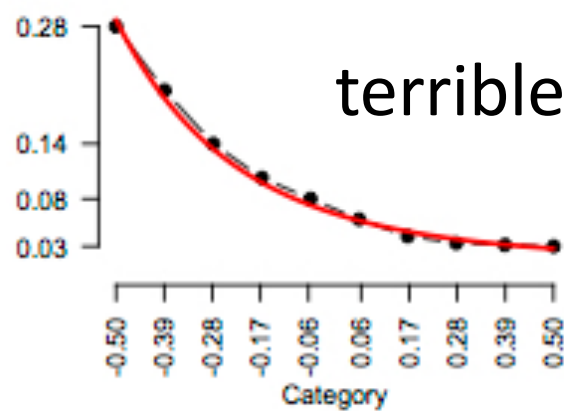
fairly



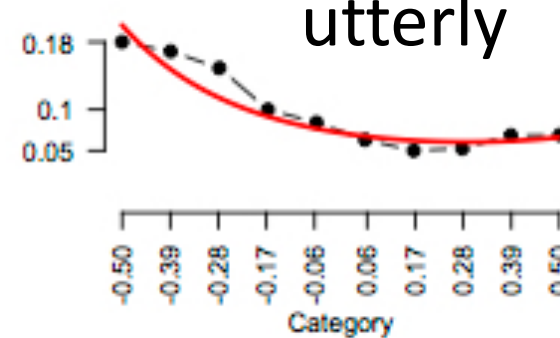
excellent



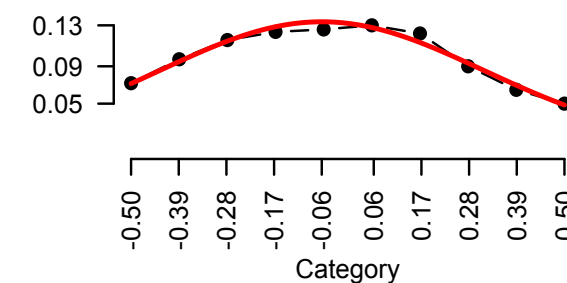
terrible



utterly



pretty



Or use regression coefficients to weight words

Train a classifier based on supervised data

- Predict: human-labeled connotation of a document
- From: all the words and bigrams in it

Use the regression coefficients as the weights

# Log odds ratio informative Dirichlet prior

Monroe, B. L., Colaresi, M. P., and Quinn, K. M. (2008). Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16(4), 372–403.

**Log likelihood ratio:** does “horrible” occur more % in corpus A or B?

$$\begin{aligned}\text{llr}(\textit{horrible}) &= \log \frac{P^i(\textit{horrible})}{P^j(\textit{horrible})} \\ &= \log P^i(\textit{horrible}) - \log P^j(\textit{horrible}) \\ &= \log \frac{f^i(\textit{horrible})}{n^i} - \log \frac{f^j(\textit{horrible})}{n^j}\end{aligned}$$

# Log odds ratio informative Dirichlet prior

Monroe, B. L., Colaresi, M. P., and Quinn, K. M. (2008). Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16(4), 372–403.

**Log odds ratio:** does “horrible” have higher odds in A or B?

$$\begin{aligned}\text{lor}(\textit{horrible}) &= \log \left( \frac{P^i(\textit{horrible})}{1 - P^i(\textit{horrible})} \right) - \log \left( \frac{P^j(\textit{horrible})}{1 - P^j(\textit{horrible})} \right) \\ &= \log \left( \frac{\frac{f^i(\textit{horrible})}{n^i}}{1 - \frac{f^i(\textit{horrible})}{n^i}} \right) - \log \left( \frac{\frac{f^j(\textit{horrible})}{n^j}}{1 - \frac{f^j(\textit{horrible})}{n^j}} \right) \\ &= \log \left( \frac{f^i(\textit{horrible})}{n^i - f^i(\textit{horrible})} \right) - \log \left( \frac{f^j(\textit{horrible})}{n^j - f^j(\textit{horrible})} \right)\end{aligned}$$

# Log odds ratio with a prior

Monroe, B. L., Colaresi, M. P., and Quinn, K. M. (2008). Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16(4), 372–403.

Log odds ratio from previous slide:

$$\log \left( \frac{f^i(\textit{horrible})}{n^i - f^i(\textit{horrible})} \right) - \log \left( \frac{f^j(\textit{horrible})}{n^j - f^j(\textit{horrible})} \right)$$

Now with a prior:

$$\delta_w^{(i-j)} = \log \left( \frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)} \right) - \log \left( \frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)} \right)$$

$n^i$  = size of corpus  $i$ ,  $n^j$  = size of corpus  $j$ ,  $f_w^i$  = count of word  $w$  in corpus  $i$ ,  $f_w^j$  = count of word  $w$  in corpus  $j$ ,  $\alpha_0$  is the size of the background corpus, and  $\alpha_w$  = count of word  $w$  in the background corpus.)



# Log odds ratio informative Dirichlet prior

Monroe, B. L., Colaresi, M. P., and Quinn, K. M. (2008). Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16(4), 372–403.

$$\sigma^2 \left( \hat{\delta}_w^{(i-j)} \right) \approx \frac{1}{y_w^i + \alpha_w} + \frac{1}{y_w^j + \alpha_w}$$

Final statistic for a word: z-score of its log-odds-ratio:

$$\frac{\hat{\delta}_w^{(i-j)}}{\sqrt{\sigma^2 \left( \hat{\delta}_w^{(i-j)} \right)}}$$

# Top 50 words associated with bad (= 1-star) reviews by Monroe, *et al.* (2008) method

Jurafsky et al., 2014

| Class                  | Words in 1-star reviews  | Class                       | Words in 5-star reviews   |
|------------------------|--|-----------------------------|---|
| <b>Negative</b>        | <i>worst, rude, terrible, horrible, bad, awful, disgusting, bland, tasteless, gross, mediocre, overpriced, worse, poor</i> | <b>Positive</b>             | <i>great, best, love(d), delicious, amazing, favorite, perfect, excellent, awesome, friendly, fantastic, fresh, wonderful, incredible, sweet, yum(my)</i> |
| <b>Negation</b>        | <i>no, not</i>   | <b>Emphatics/universals</b> | <i>very, highly, perfectly, definitely, absolutely, everything, every, always</i>   |
| <b>1Pl pro</b>         | <i>we, us, our</i>   | <b>2 pro</b>                | <i>you</i>  |
| <b>3 pro</b>           | <i>she, he, her, him</i>   | <b>Articles</b>             | <i>a, the</i>   |
| <b>Past verb</b>       | <i>was, were, asked, told, said, did, charged, waited, left, took</i>  | <b>Advice</b>               | <i>try, recommend</i>   |
| <b>Sequencers</b>      | <i>after, then</i>   | <b>Conjunct</b>             | <i>also, as, well, with, and</i>  |
| <b>Nouns</b>           | <i>manager, waitress, waiter, customer, customers, attitude, waste, poisoning, money, bill, minutes</i>                    | <b>Nouns</b>                | <i>atmosphere, dessert, chocolate, wine, course, menu</i>   |
| <b>Irrealis modals</b> | <i>would, should</i>   | <b>Auxiliaries</b>          | <i>is/'s, can, 've, are</i>   |
| <b>Comp</b>            | <i>to, that</i>  | <b>Prep, other</b>          | <i>in, of, die, city, mouth</i>   |

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Supervised Learning of  
Word Sentiment

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Using the lexicons to  
detect affect

# Lexicons for detecting document affect: Simplest unsupervised method

## Sentiment:

- Sum the weights of each positive word in the document
- Sum the weights of each negative word in the document
- Choose whichever value (positive or negative) has higher sum

## Emotion:

- Do the same for each emotion lexicon

Lexicons for detecting document affect:  
Simplest unsupervised method

$$f^+ = \sum_{w \text{ s.t. } w \in \text{positivelexicon}} \theta_w^+ \text{count}(w)$$

$$f^- = \sum_{w \text{ s.t. } w \in \text{negativelexicon}} \theta_w^- \text{count}(w)$$

Sentiment = + if  $f^+ > f^-$

Lexicons for detecting document affect:  
Slightly more complex unsupervised method

$$f^+ = \sum_{w \text{ s.t. } w \in \text{positivelexicon}} \theta_w^+ \text{count}(w)$$

$$f^- = \sum_{w \text{ s.t. } w \in \text{negativelexicon}} \theta_w^- \text{count}(w)$$

$$\text{sentiment} = \begin{cases} + & \text{if } \frac{f^+}{f^-} > \lambda \\ - & \text{if } \frac{f^-}{f^+} > \lambda \\ 0 & \text{otherwise.} \end{cases}$$

# Lexicons for detecting document affect: Simplest supervised method

Use the lexicons as **features** for a classifier

Given a training set

- Each observation has a label (review X has sentiment Y)
- Assign features to each observation
- Use “counts of lexicon categories” as a features
  - NRC Emotion category “anticipation” had count of 2
    - 2 words in this document were in “anticipation” lexicon
  - LIWC category “cognition” had count of 7



# Lexicons for detecting document affect: Simplest supervised method

## Baseline

- Use counts of **all** the words and bigrams in the training set
  - Like the naïve bayes algorithm
- **This is hard to beat**
- But “using all the words” only works if the training and test sets are very similar
- In real life, sometimes the test set is very different
  - Lexicons are useful in that situation

# Computing entity-centric affect

Suppose we want an affect score for an entity in a text (not the entire document)

Entity-centric method of Field and Tsvetkov (2019)

# Entity-centric affect (Field and Tsvetkov 2019)

1: Train classifier to predict V/A/D from embeddings

1. For each word  $w$  in the training corpus:
  - Use off-the-shelf encoders (like BERT) to extract a contextual embedding  $e$  for each instance of the word.
  - Average over the  $e$  embeddings of each instance of  $w$  to obtain a single embedding vector for one training point  $w$ .
  - Use the NRC Lexicon to get V, A, and D scores for  $w$ .
2. Train (three) regression models on all words  $w$  to predict V, A, D scores from a word's average embedding.

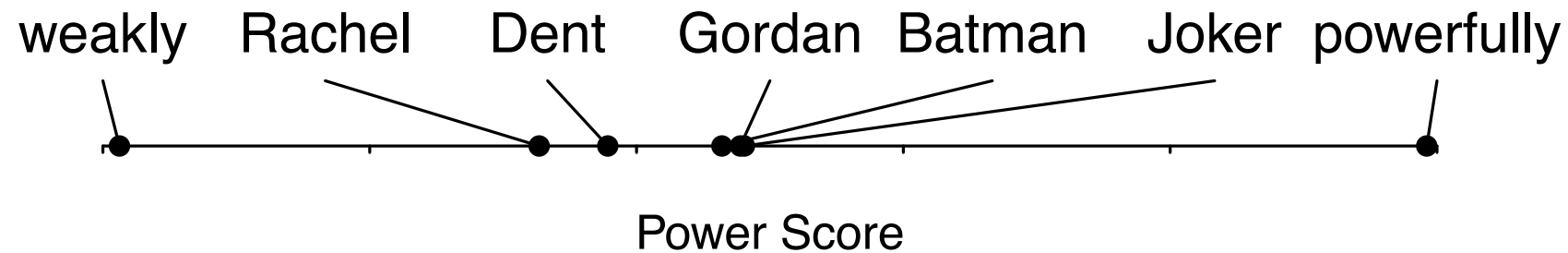
# Entity-centric affect (Field and Tsvetkov 2019)

2: Assign scores to entity mentions

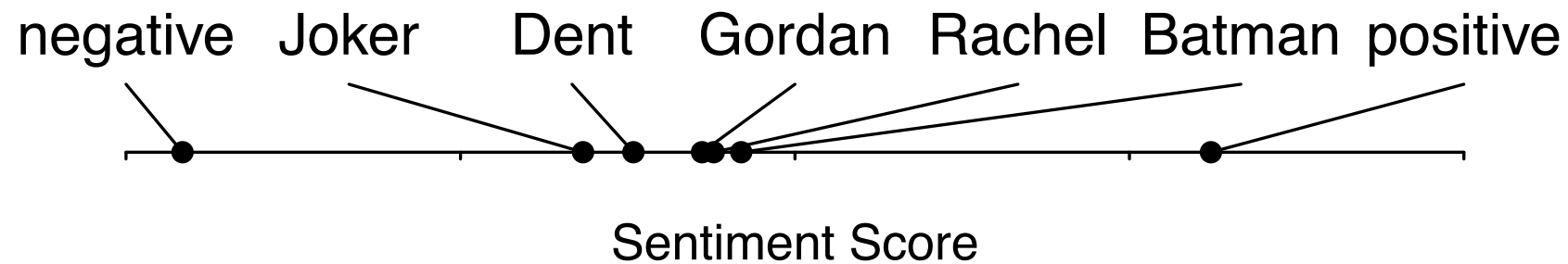
Given an entity mention  $m$  in a text, assign affect scores as follows:

- Use the same pretrained LM to get contextual embeddings for  $m$  in context.
- Feed this embedding through the 3 regression models to get V, A, D scores for the entity.

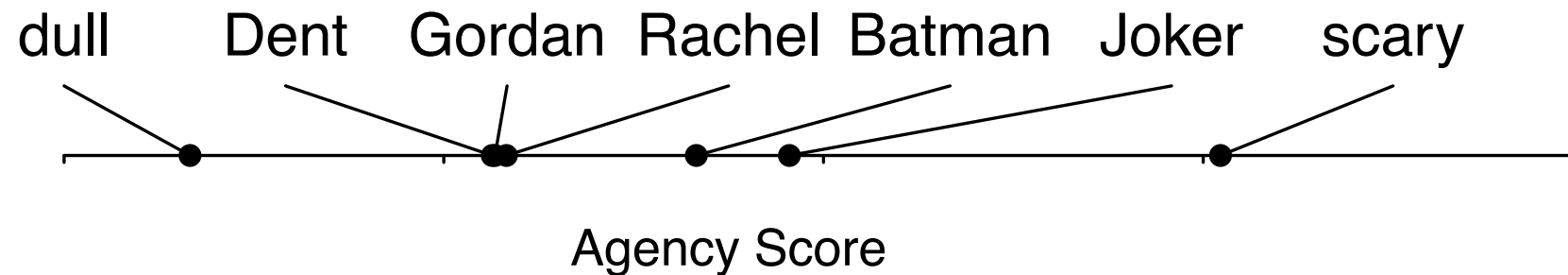
# Entity-centric affect (Field and Tsvetkov 2019)



= dominance



= valence



= arousal

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Using the lexicons to  
detect affect

Lexicons for  
Sentiment,  
Affect, and  
Connotation

Connotation frames

# Connotation Frames intuition

A predicate expresses connotations about its arguments (Rashkin et al. 2016, Rashkin et al. 2017).

By using *violate*, author is sympathizing with B, and expressing negative sentiment toward A:

- Country A violated the sovereignty of Country B

By using *survive*, author is saying that the bombing is negative, and sympathizing with teenager:

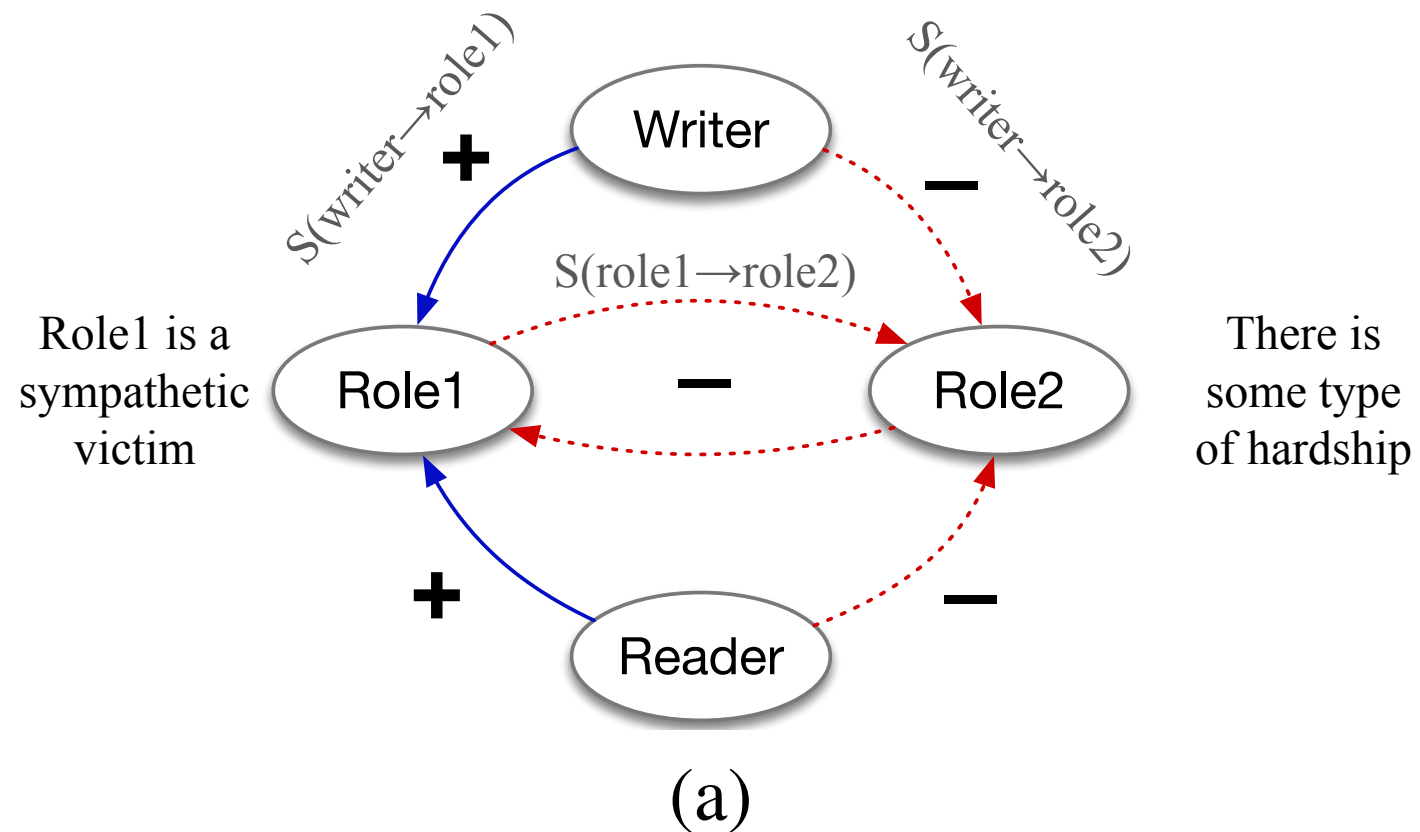
- the teenager ... survived the Boston Marathon bombing”



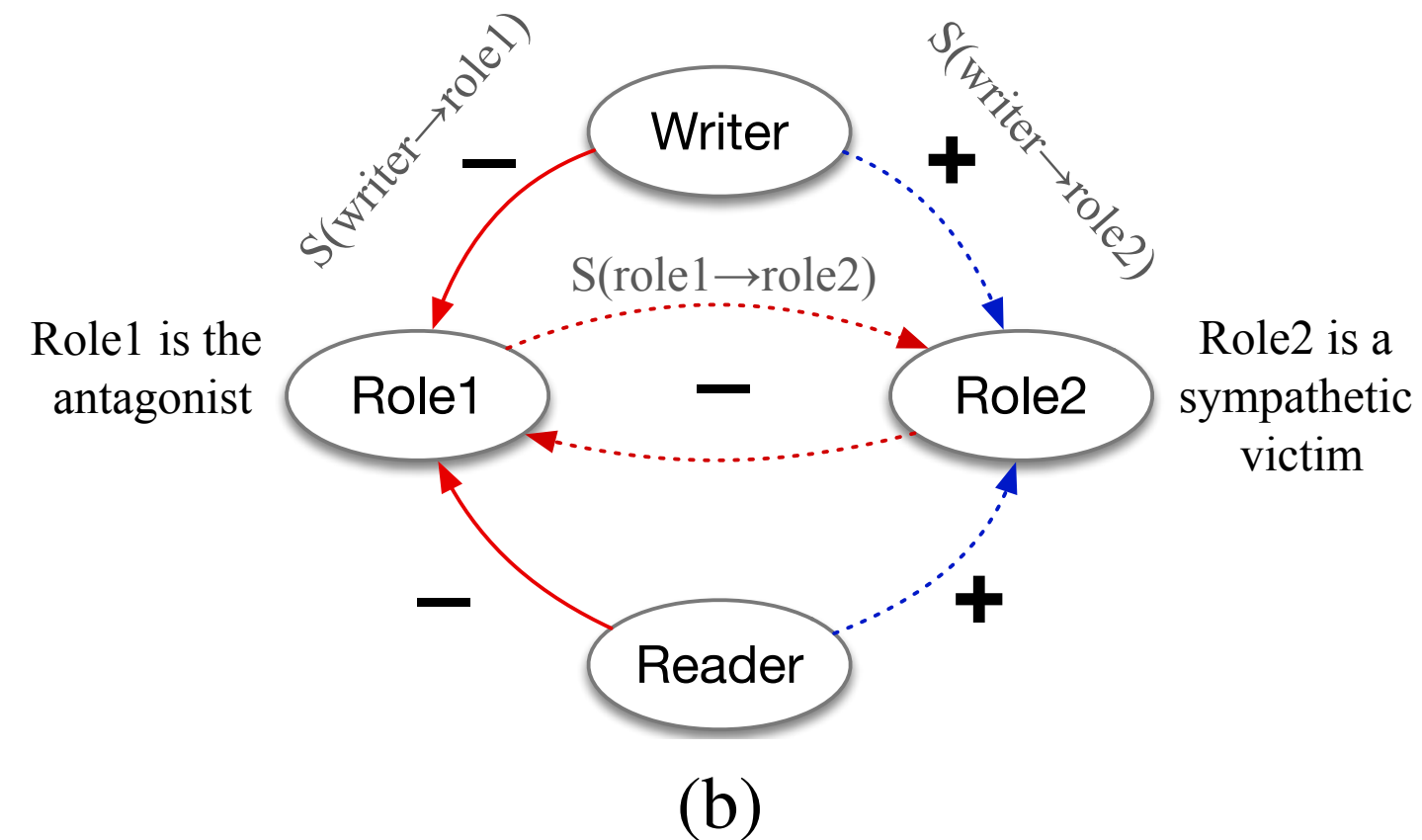
# Connotation Frames

Rashkin et al., 2016, 2017

Connotation Frame for “Role1 *survives* Role2”



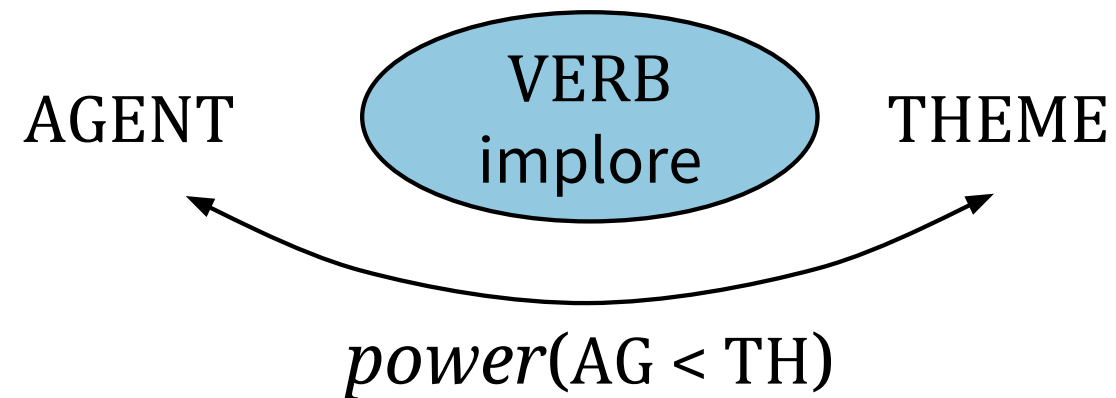
Connotation Frame for “Role1 *violates* Role2”



# Connotation Frames can also mark power and agency

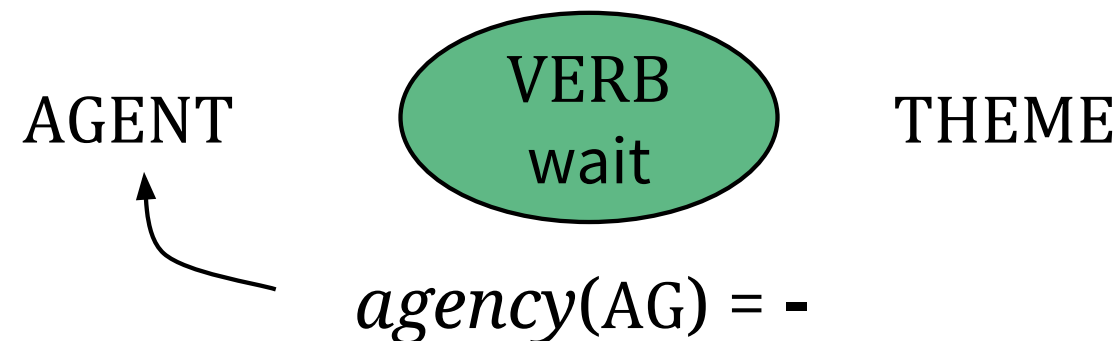
Sap et al. 2017

He **implored** the tribunal to show mercy.



---

The princess **waited** for her prince.



Lexicons for  
Sentiment,  
Affect, and  
Connotation

Connotation frames