CS224C: NLP for CSS

Sentiment and Affect

Diyi Yang

Stanford CS

Lecture Overview

- Emotion
- Subjectivity
- LIWC
- Empath
- Semi-supervised and supervised approaches to infer affect

Some slides are adapted based on *Lexicons for Sentiment, Affect, and Connotation* from Speech and Language Processing (3rd ed. draft) Dan Jurafsky and James H. Martin (https://web.stanford.edu/~jurafsky/slp3/)

Lexicon

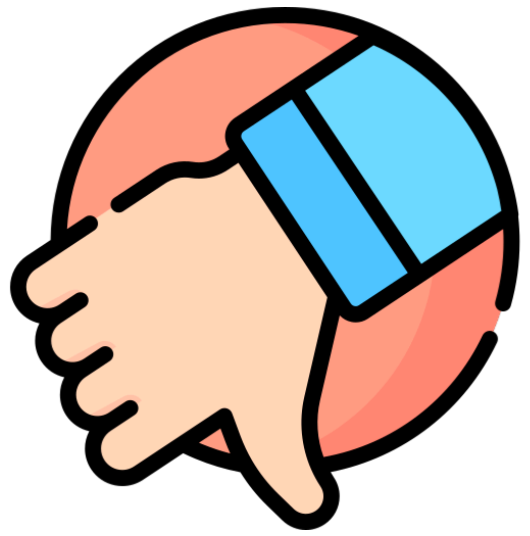
- A (usually hand-built) list of words that correspond to some meaning or class
- Possibly with numeric values
- Commonly used as simple classifiers, or as features to complex classifiers

Why Lexicons for Sentiment and Affect

Easy to use

Interpretable

Fast to calculate



Fail to consider negation or word order

Can't deal with context

Scherer's typology of affective states

Emotion: brief organically synchronized evaluation of a major event

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

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

cheerful, gloomy, irritable, listless, depressed, buoyant

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

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, 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

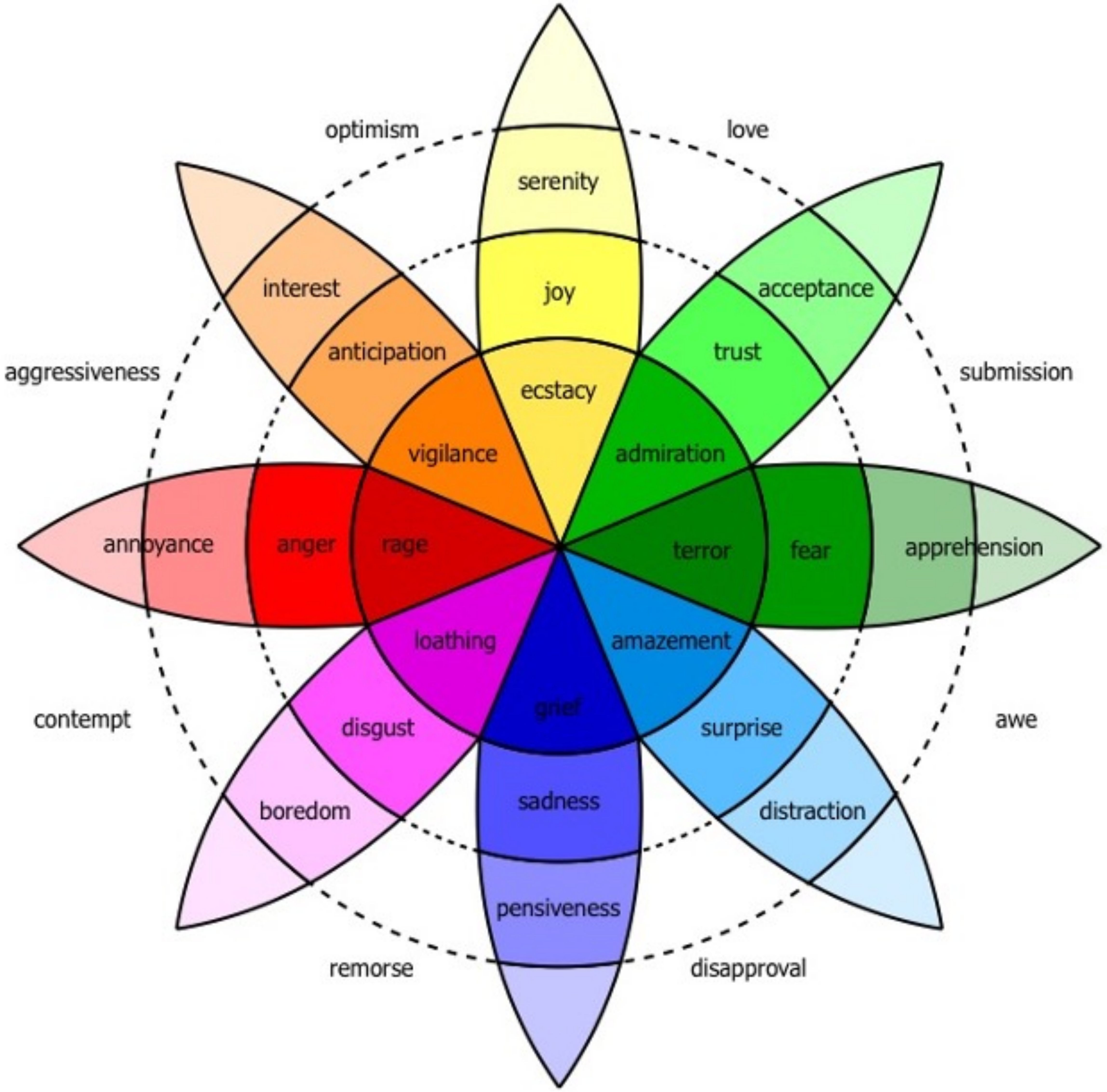
- Valence (positive negative)
- Arousal (strong, weak)
- Control

Ekman’s 6 basic emotions:
Surprise, happiness, anger, fear, disgust, sadness

Plutchick’s wheel of emotion

8 basic emotions

In four opposing pairs:
- joy–sadness
- anger–fear
- trust–disgust
- anticipation–surprise



Alternative: spatial model

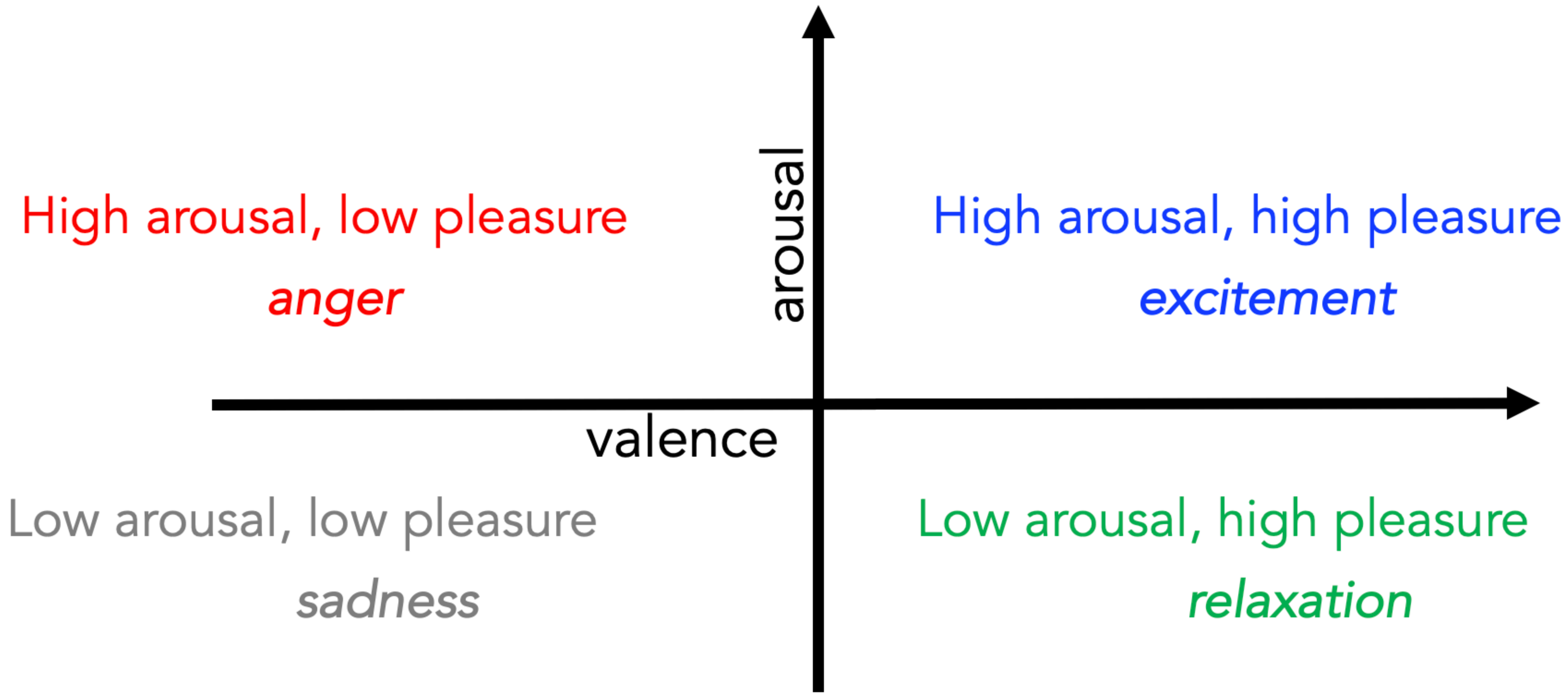
An emotion is a point in 2- or 3-dimensional space

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



Some Sentiment Lexicons

The General Inquirer

Positive (1915 words), and Negative (2291 words)

MPQA Subjectivity Cues Lexicon

6885 words on strong/weak subjectivity

Is a subjective word positive or negative?

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

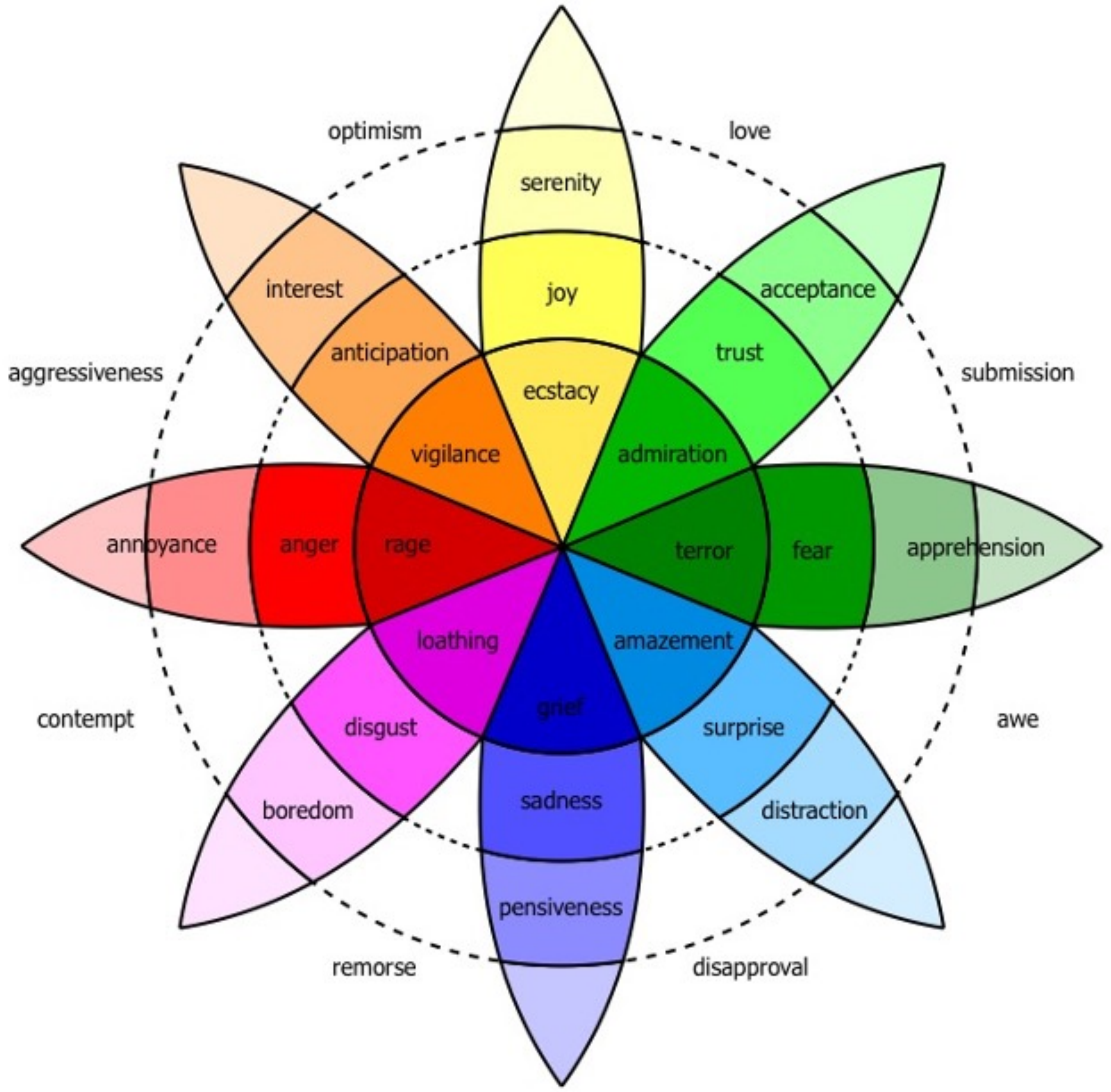
Words with consistent sentiment across lexicons

Positive	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
Negative	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

NRC Emotion Lexicon

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
```



NRC Emotion/Affect Intensity Lexicon

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

Another Widely Used Lexicon: LIWC

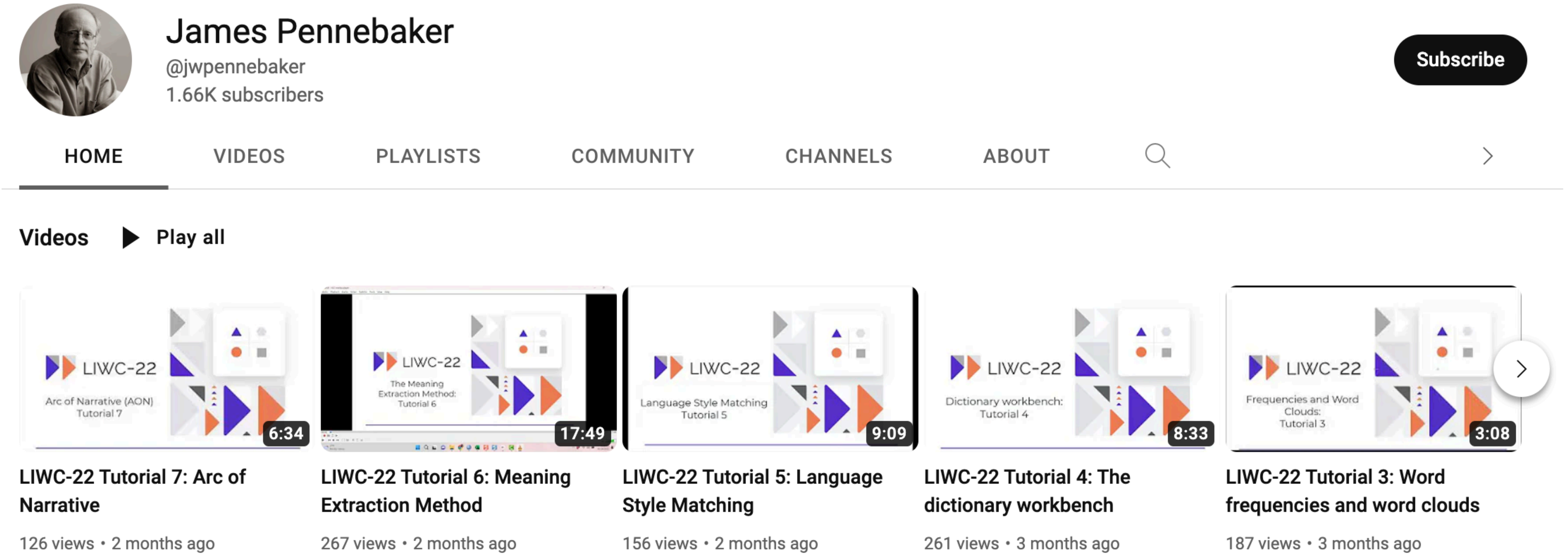
LIWC: Linguistic Inquiry and Word Count

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

http://www.liwc.net/
2300 words
>70 classes

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

LIWC: Linguistic Inquiry and Word Count



Concreteness versus Abstractness

Definition:

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

Lexicon:

37,058 English words and 2,896 two-word expressions

Rating from 1 (abstract) to 5 (concrete)

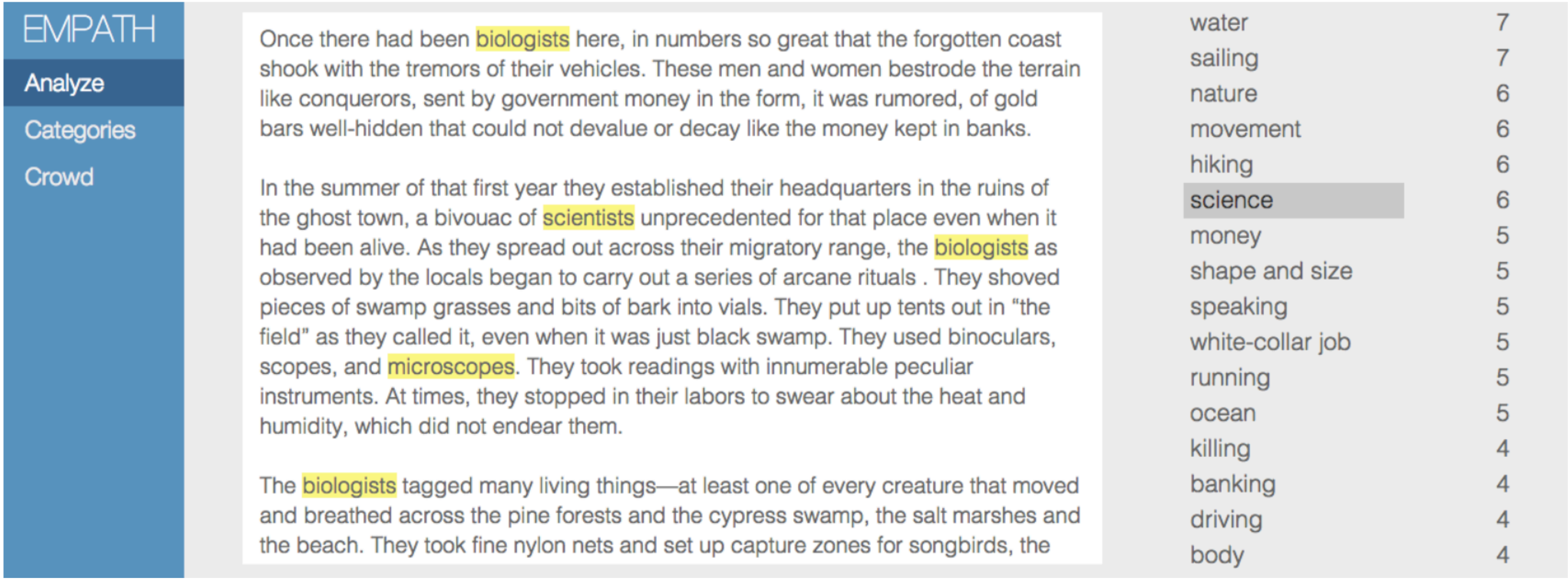
Brysbaert, M., Warriner, A. B., and Kuperman, V. (2014) Concreteness ratings for 40 thousand generally known English word lemmas Behavior Research Methods.

Concreteness versus Abstractness

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
```

Empath



Fast, Ethan, Binbin Chen, and Michael S. Bernstein. "Empath: Understanding topic signals in large-scale text." In Proceedings of the 2016 CHI conference on human factors in computing systems, pp. 4647-4657. 2016.

Empath

Generate categories from seed words using word embeddings

Broad set of 200 built-in categories:

```
Technology = {iPad, android, …}
Violence = {bleed, punch, …}
Government = {embassy, democrat, …}
```

When someone is punching the printer in the computer lab because of a paper jam.

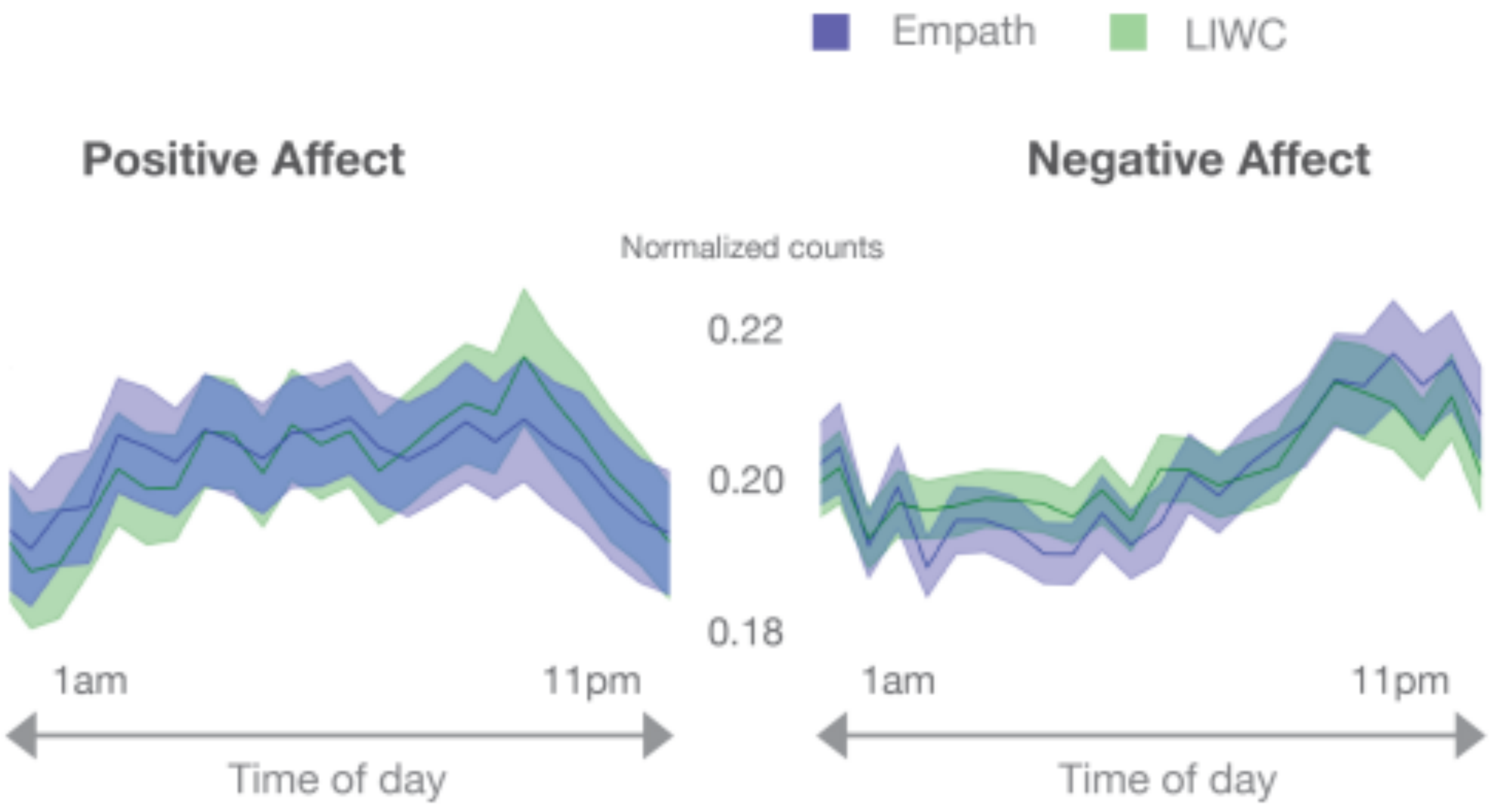
I'm scared to learn cause I'm scared of truth.

Spending days off class to online chat with course support. Help me Adeep!

Now Russia is the nation going thru the kicking out the ruling party drama.

social media	war	violence	technology	fear	pain	hipster	contempt
facebook	attack	hurt	ipad	horror	hurt	vintage	disdain
instagram	battlefield	break	internet	paralyze	pounding	trendy	mockery
notification	soldier	bleed	download	dread	sobbing	fashion	grudging
selfie	troop	broken	wireless	scared	gasp	designer	haughty
account	army	scar	computer	tremor	torment	artsy	caustic
timeline	enemy	hurting	email	despair	groan	1950s	censure
follower	civilian	injury	virus	panic	stung	edgy	sneer

Empath correlates with LIWC well



Lexicon based computing for sentiment/affect

Ratio of words in a sentence belonging to a category

$$r_k = \frac{\mathbf{n}_k}{\mathbf{n}}$$

The number of words in a given sentence belonging to a category k

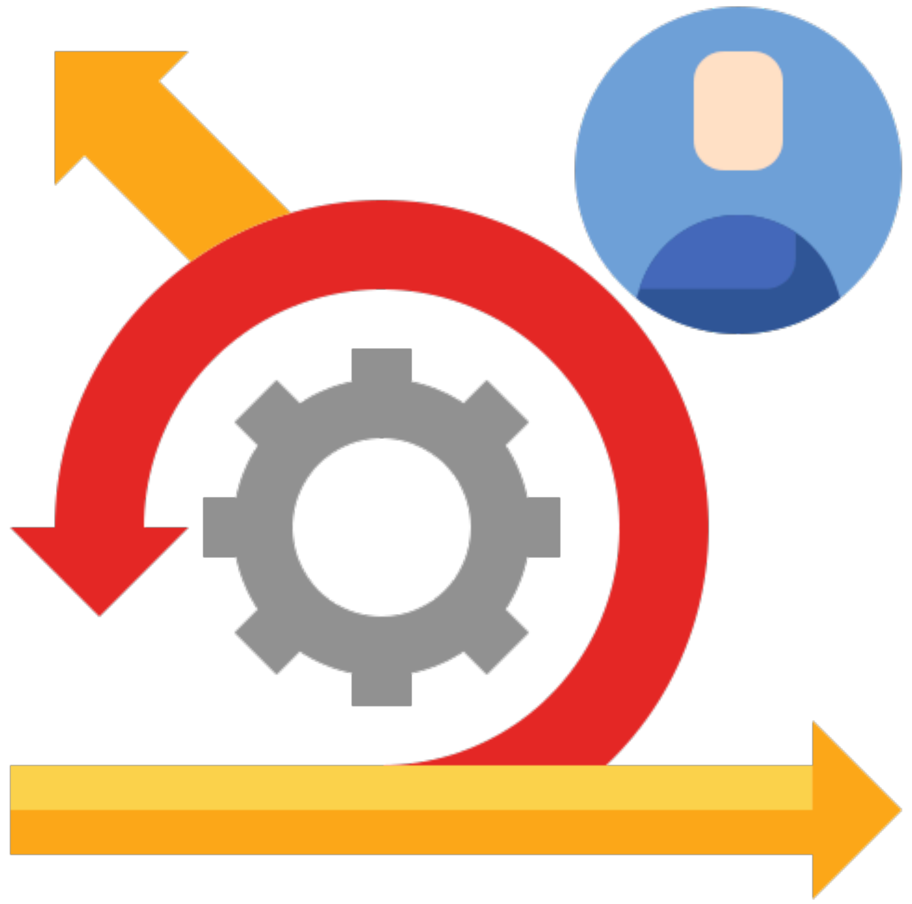
The total number of words in a given sentence

So far, only lexicon based approaches …

Supervised approaches exist

Or building lexicons via human annotation

Or semi-supervised induction



Semantic Axis Methods

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

For each word to be added to lexicon

1. Compute a word representation
2. Use this to measure its distance from the poles
3. Assign it to the pole it is closer to

Initial Seeds for Different Domains

- Start with a single large seed lexicon and rely on the induction algorithm to fine-tune it to the domain
- 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

Computing word 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

Representing each pole

Start with embeddings for seed words:

$$S^+ = \{E(w_1^+), E(w_2^+), ..., E(w_n^+)\}$$
$$S^- = \{E(w_1^-), E(w_2^-), ..., E(w_m^-)\}$$

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:

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

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

Word score is cosine with axis

$$\text{score}(w) = \cos\left(E(w), \mathbf{V}_{\text{axis}}\right) = \frac{E(w) \cdot \mathbf{V}_{\text{axis}}}{\|E(w)\|\|\mathbf{V}_{\text{axis}}\|}$$

Supervised Learning of Word Sentiment

Learn word sentiment supervised by online review scores

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)

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

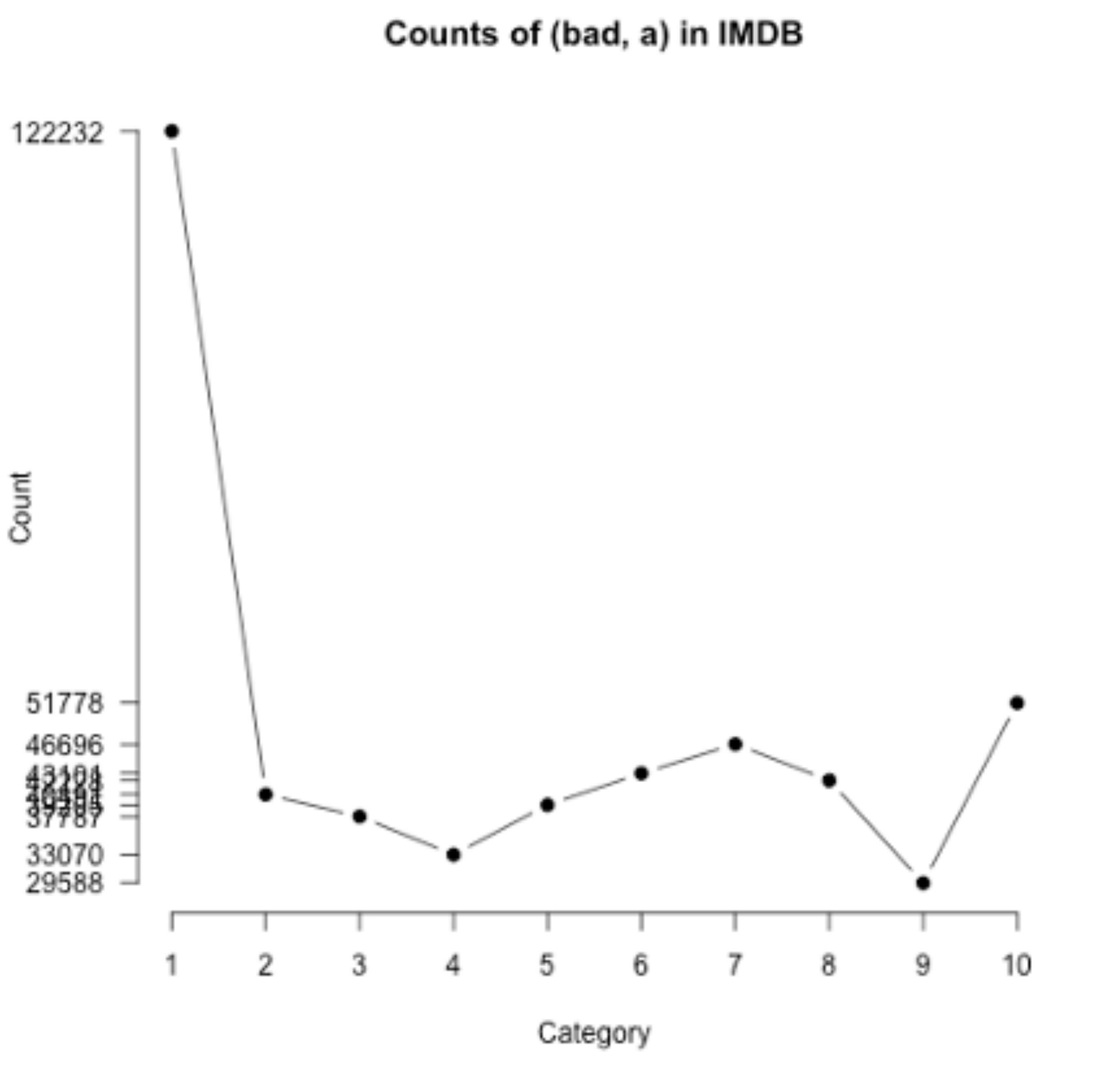
Online Review Data

	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

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

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



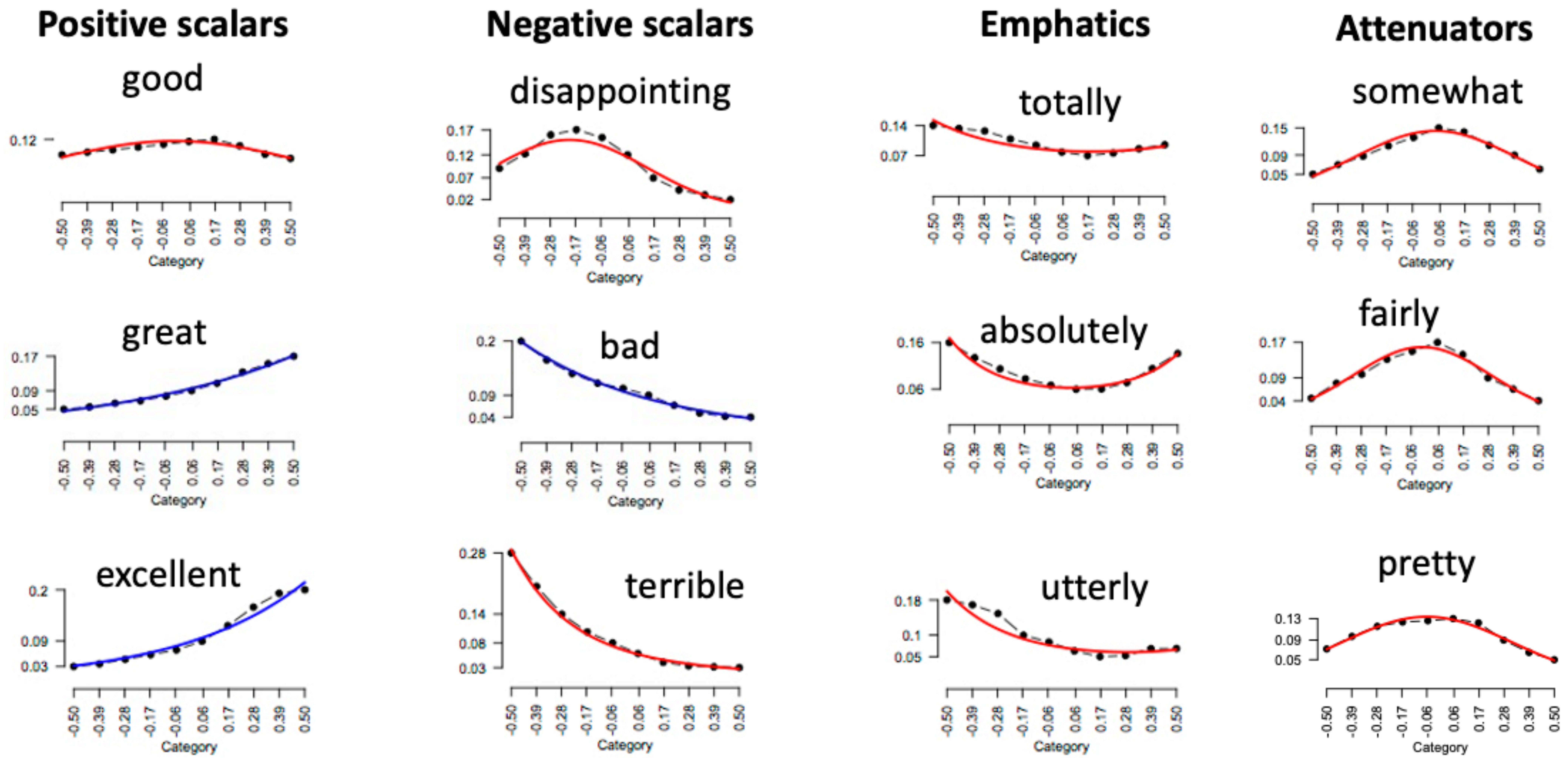
But can't use raw counts; instead, likelihood:

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

Make them comparable between words via Scaled likelihood: $\dfrac{P(w|c)}{P(w)}$

"Potts diagrams"

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



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

Log likelihood ratio: does "horrible" occur more % in corpus i or j?

llr

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

Log odds ratio

Log odds ratio: does “horrible” have a higher odds in corpus i or j?

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

Log odds ratio with a prior

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

Now with prior

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

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

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, α_0 is the size of the background corpus, and α_w = count of word w in the background corpus.)

Log odds ratio with a Dirichlet prior

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

Final statistic for a word: z-score of its log-odds-ratio: $$\frac{\hat{\delta}_w^{(i-j)}}{\sqrt{\sigma^2\left(\hat{\delta}_w^{(i-j)}\right)}}$$

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.

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

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

Jurafsky, D., V. Chahuneau, B. R. Routledge, and N. A. Smith. 2014. Narrative framing of consumer sentiment in online restaurant reviews. First Monday, 19(4).

Summary

- ✓ Emotion can be represented by fixed atomic units often called basic emotions, or as points in space defined by dimensions like valence and arousal.
- ✓ Affective lexicons can be built by hand, using crowd sourcing to label the affective content of each word.
- ✓ Lexicons can be built with semi-supervised, bootstrapping from seed words using similarity metrics like embedding cosine.
- ✓ Lexicons can be learned in a fully supervised manner, when a convenient training signal can be found in the world, such as ratings assigned by users on a review site.
- ✓ Words can be assigned weights in a lexicon by using various functions of word counts, and ratio metrics like log odds ratio informative Dirichlet prior