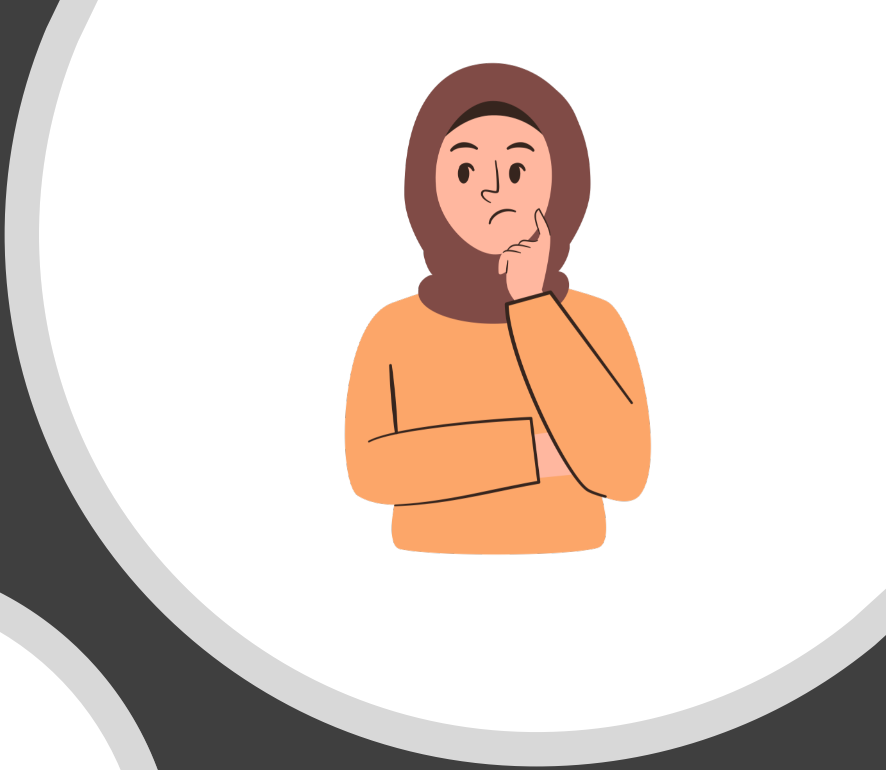
Affective Lexica and Referring Expressions

Natalie Parde

UIC CS 421

Language is tricky for many reasons!

- Ambiguity
- Abstractness
- Tone
- Polarity
- Subjectivity



To understand meaning and intent, we often have to read between the lines.

- Difficult for people, and extremely challenging for machines!
- Approaches we've examined so far that help us with this:
  - Information extraction
  - Word sense disambiguation
- Another helpful tool:
  - **Affective analysis**

This Week's Topics

Tuesday

- Affective Analysis
- Affective Lexicons
- Inducing Affect and Association

Thursday

- Affective Tasks
- Coreference Resolution
- Referring Expressions

This Week's Topics

Affective Analysis
Affective Lexicons
Inducing Affect and Association

Tuesday

Thursday

Affective Tasks
Coreference Resolution
Referring Expressions

Affective Analysis

- The automated analysis of the **emotions**, **moods**, **attitudes**, **stance**, or **personality** that is conveyed or evoked by a language sample
- Popular tool to facilitate social science research
  - Determining views towards a specific topic
  - Assessing public opinion
  - Interpreting intent

When can these different forms of analysis be useful?

- **Attitudes** help us figure out what people like or dislike
  - Useful for processing and interpreting reviews
  - Useful for measuring public sentiment
- **Emotions** and **moods** help us measure engagement or frustration, among other factors
  - Useful for studying how people interact with automated systems
  - Useful for psycholinguistic tasks

When can these different forms of analysis be useful?

- **Interpersonal stance** can help us understand perspectives and characteristics of multi-party interaction
  - Useful for determining views with respect to specific topics
  - Useful for summarizing conversations (is the interaction friendly or awkward?)
- **Personality** can help us customize interactive agents
  - Useful for matching user expectations
  - Useful for optimizing user experience

These analyses give rise to many interesting NLP tasks!

Attitude prediction and role framing:

- https://aclanthology.org/W17-6917.pdf

Emotion classification:

- https://aclanthology.org/S19-2005/
- https://paperswithcode.com/task/emotion-classification

Mood tracking:

- https://aclanthology.org/2022.clpsych-1.22/

Stance detection:

- https://paperswithcode.com/task/stance-detection
- https://dl.acm.org/doi/abs/10.1145/3369026

Personality prediction:

- https://aclanthology.org/2023.wassa-1.34.pdf

Automated Affect Recognition

- Classification of text into predetermined affective categories
- Commonly performed using supervised learning
- Useful features:
  - N-grams
  - Features derived from **affective lexica**

This Week's Topics

Tuesday

- Affective Analysis
- Affective Lexicons
- Inducing Affect and Association

Thursday

- Affective Tasks
- Coreference Resolution
- Referring Expressions

What is an affective lexicon?

- Known list of words corresponding to different affective dimensions
  - Optionally including scores indicating the closeness of their association with a given dimension
- This information makes it easier to predict affect for an entire instance
  - If the input contains many positive words, it is likely a positive input!

What are some common forms of affective analysis?

SENTIMENT ANALYSIS

EMOTION RECOGNITION

CONNOTATION FRAMING

Emotion Recognition

- Emotion can be defined in numerous ways
- In some frameworks, emotion is an **atomic unit**
  - IsHappy = TRUE
- In other frameworks, emotion is a **point along a multi-dimensional continuum**
  - Happiness = 0.78

Common Emotion Frameworks

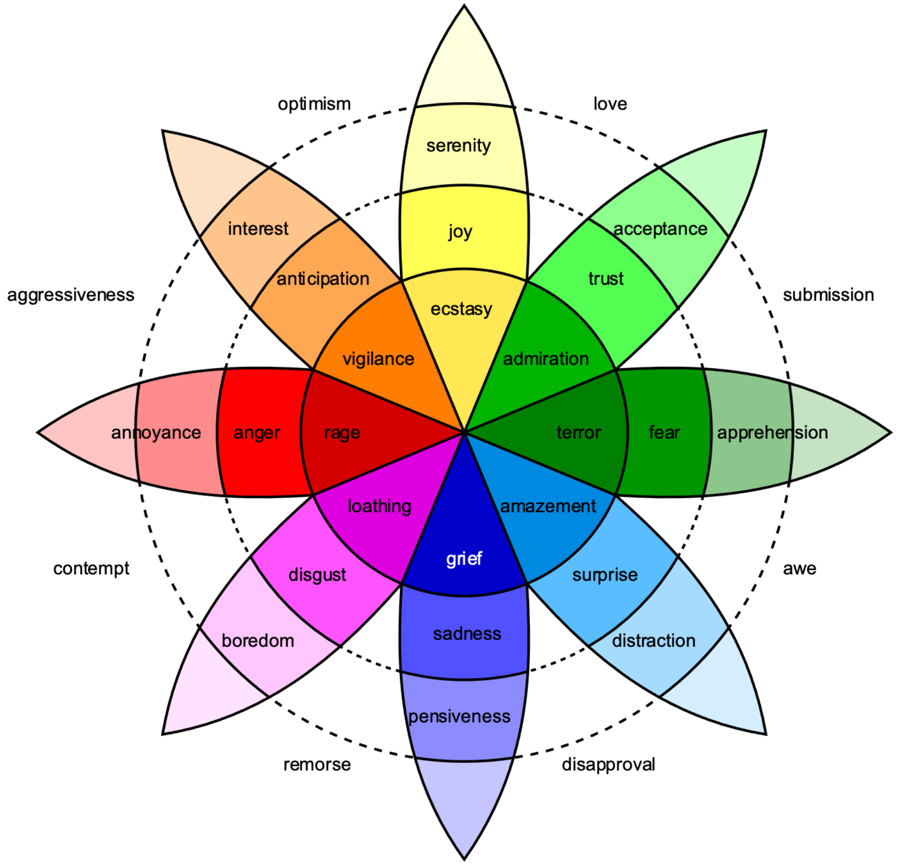
- Ekman’s six basic emotions
- Plutchik’s wheel of emotion

Ekman's Basic Emotions

- Paper:
  - Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. J. Power (Eds.), Handbook of cognition and emotion (pp. 45–60). John Wiley & Sons Ltd. https://doi.org/10.1002/0470013494.ch3
- Six basic emotions:
  - Happiness
  - Sadness
  - Anger
  - Fear
  - Disgust
  - Surprise
- Emotions are distinct from one another
- Generally known to be present across cultures

Plutchik Wheel of Emotion

- Paper:
  - Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *Theories of emotion* (pp. 3-33). Academic press. https://doi.org/10.1016/B978-0-12-558701-3.50007-7
- Situates eight basic emotions in a wheel
- Emotions located opposite one another also oppose one another semantically
- Stronger emotions derived from the basic emotions are located at more internal locations
- Weaker emotions derived from the basic emotions are located at more external locations



Atomic vs. Continuous Emotions

- Ekman and Plutchik both define emotion as an atomic unit
- Emotion along a continuum is often represented using a set of common dimensions
  - **Valence:** Pleasantness (e.g., positive or negative)
  - **Arousal:** Intensity of emotion provoked (e.g., strong or weak)
  - (Sometimes) **Dominance:** Degree of control exerted (e.g., active or passive)
- Sentiment is sometimes viewed as a measure of valence

Sentiment and Affect Lexicons

- A wide range of resources for sentiment and affect recognition are available for public use!
- Can be highly useful for performing automated sentiment or affect analysis

General Inquirer

- Classic resource created during the 1960s
  - Stone, P. J., In Kirsch, J., & Cambridge Computer Associates. (1966). *The general inquirer: A computer approach to content analysis*. Cambridge, Mass: M.I.T. Press.
- 1915 positive words
- 2291 negative words
- Additional words associated with other categories
- Information:
  - https://inquirer.sites.fas.harvard.edu/
  - https://web.archive.org/web/20110805192759/http://www.webuse.umd.edu:9090/

MPQA Subjectivity Lexicon

- Collection of positive and negative words from existing lexicons
  - 2718 positive words
  - 4912 negative words
- Additional subjective words learned via bootstrapping, with manually-provided sentiment and subjectivity levels
  - Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 347-354).
- Link:
  - https://mpqa.cs.pitt.edu/lexicons/subj_lexicon/

Opinion Lexicon

- Positive and negative words collected from product reviews via bootstrapping
  - Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 168-177).
- 2006 positive words
- 4783 negative words
- Link:
  - https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

NRC Valence, Arousal, and Dominance Lexicon

- 20,000 words labeled with valence, arousal, and dominance scores
  - Mohammad, S. (2018a). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 174-184).
- Link:
  - https://saifmohammad.com/WebPages/nrc-vad.html

VAD Lexicon Scores

- Translations are available in 100+ languages

The NRC VAD Lexicon

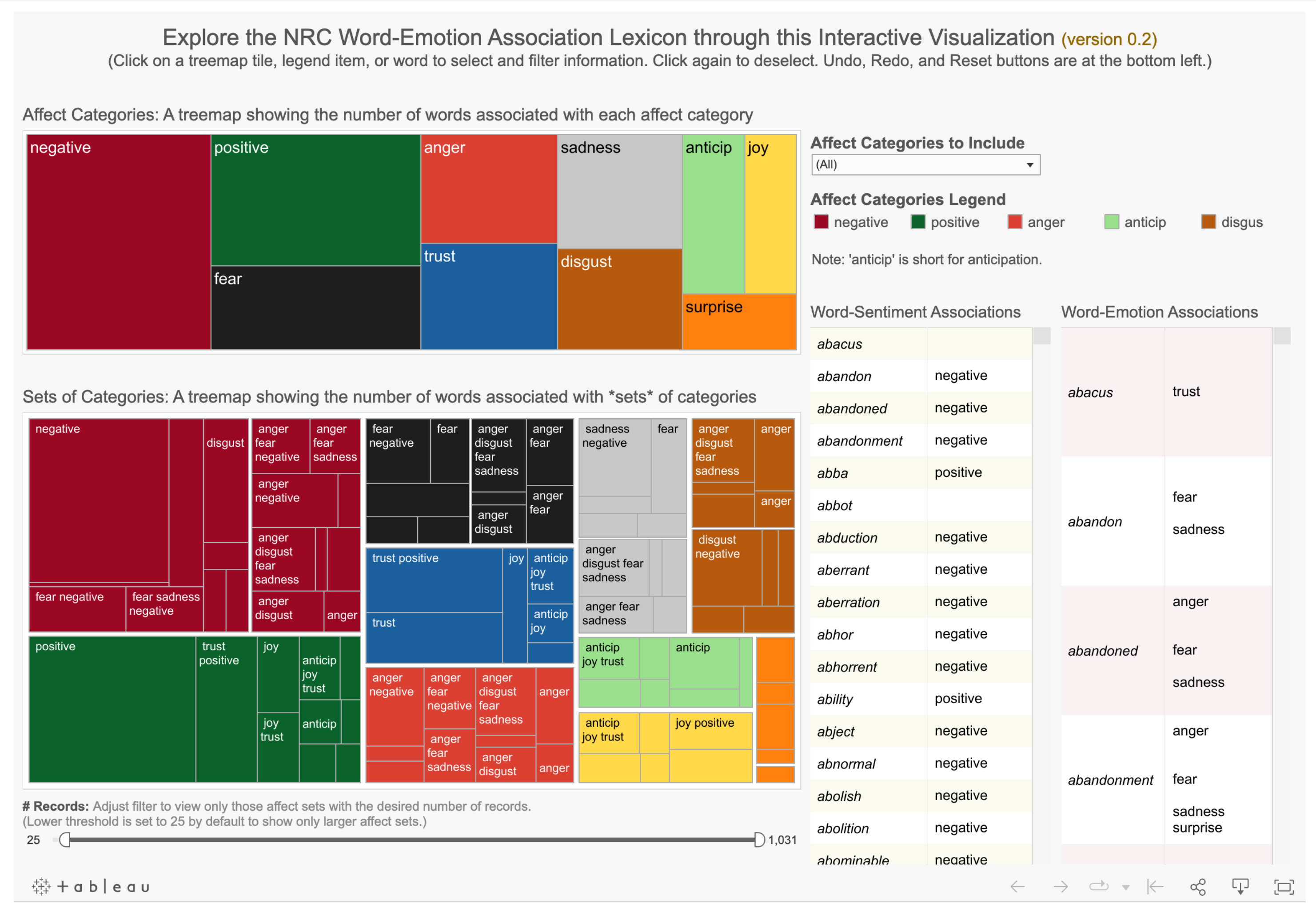
term	valence	arousal	dominance
aaaaaaah	0.479	0.606	0.291
aaaah	0.520	0.636	0.282
aardvark	0.427	0.490	0.437
aback	0.385	0.407	0.288
abacus	0.510	0.276	0.485
abalone	0.500	0.480	0.412
abandon	0.052	0.519	0.245
abandoned	0.046	0.481	0.130
abandonment	0.128	0.430	0.202
abashed	0.177	0.644	0.307
abate	0.255	0.696	0.604
abatement	0.388	0.338	0.336

Most extreme scores for each dimension:

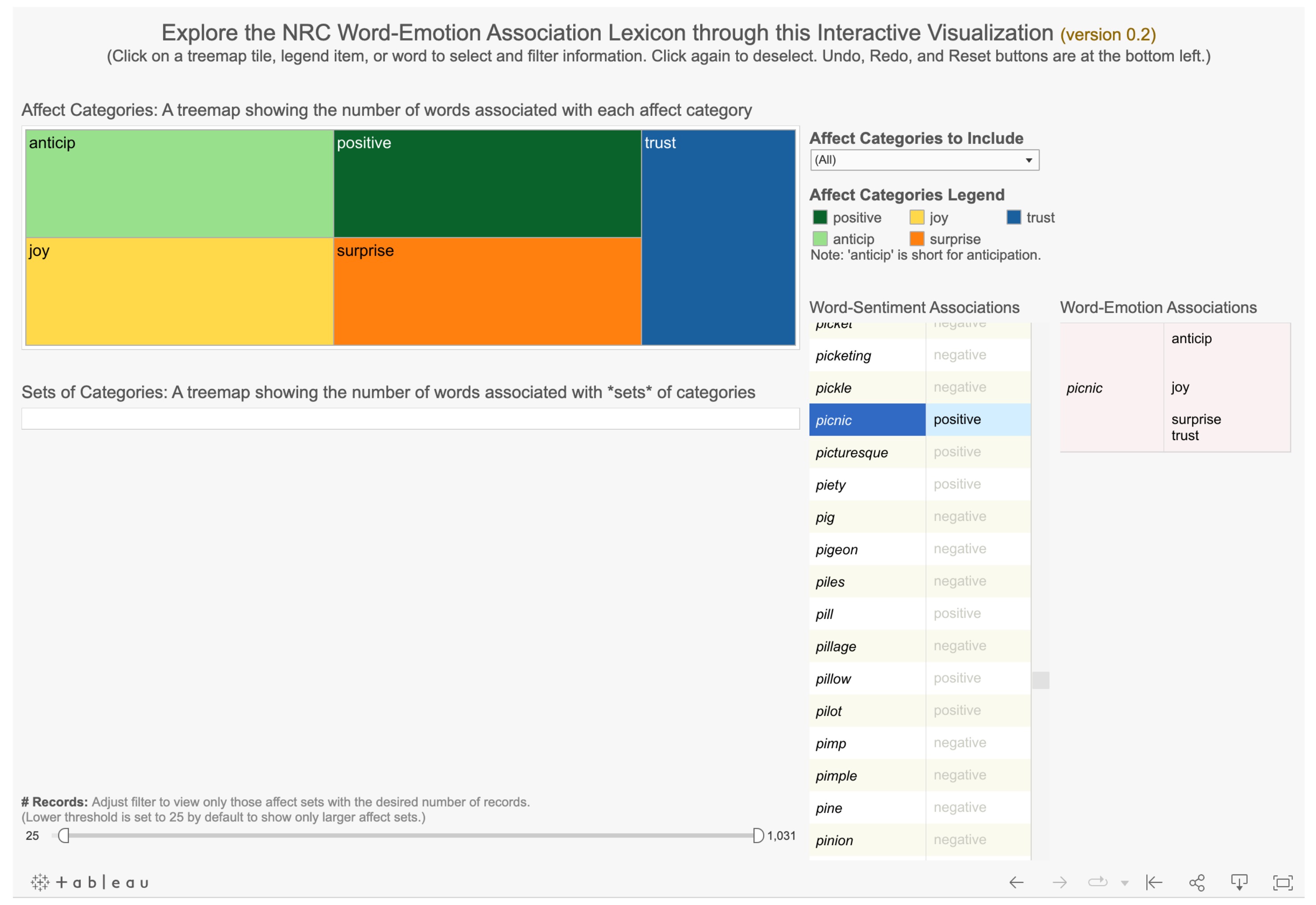
Dimension	Word	Score ⬆	Word	Score ⬇
valence	love	1.000	toxic	0.008
arousal	abduction	0.990	mellow	0.069
dominance	powerful	0.991	empty	0.081

NRC Word-Emotion Association Lexicon

- Approximately 14,000 words labeled for the eight basic emotions from Plutchik's wheel of emotions
  - Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3), 436-465.
- Link:
  - https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm



Interactive Visualization



Interactive Visualization

NRC Emotion Intensity Lexicon

- Approximately 10,000 words labeled with continuous scores for the eight basic emotions from Plutchik's wheel of emotions
  - Mohammad, S. (2018b). Word Affect Intensities. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- Link:
  - https://www.saifmohammad.com/WebPages/AffectIntensity.htm

Word	Anger	Word	Fear	Word	Joy	Word	Sadness
outraged	0.964	*horror*	0.923	*sohappy*	0.868	*sad*	0.844
brutality	0.959	*horrified*	0.922	*superb*	0.864	*suffering*	0.844
satanic	0.828	*hellish*	0.828	*cheered*	0.773	*guilt*	0.750
hate	0.828	*grenade*	0.828	*positivity*	0.773	*incest*	0.750
violence	0.742	*strangle*	0.750	*merrychristmas*	0.712	*accursed*	0.697
molestation	0.742	*tragedies*	0.750	*bestfeeling*	0.712	*widow*	0.697
volatility	0.687	*anguish*	0.703	*complement*	0.647	*infertility*	0.641
eradication	0.685	*grisly*	0.703	*affection*	0.647	*drown*	0.641
cheat	0.630	*cutthroat*	0.664	*exalted*	0.591	*crumbling*	0.594
agitated	0.630	*pandemic*	0.664	*woot*	0.588	*deportation*	0.594
defiant	0.578	*smuggler*	0.625	*money*	0.531	*isolated*	0.547
coup	0.578	*pestilence*	0.625	*rainbow*	0.531	*unkind*	0.547
overbearing	0.547	*convict*	0.594	*health*	0.493	*chronic*	0.500
deceive	0.547	*rot*	0.594	*liberty*	0.486	*injurious*	0.500
unleash	0.515	*turbulence*	0.562	*present*	0.441	*memorials*	0.453
bile	0.515	*grave*	0.562	*tender*	0.441	*surrender*	0.453
suspicious	0.484	*failing*	0.531	*warms*	0.391	*beggar*	0.422
oust	0.484	*stressed*	0.531	*gesture*	0.387	*difficulties*	0.421
ultimatum	0.439	*disgusting*	0.484	*healing*	0.328	*perpetrator*	0.359
deleterious	0.438	*hallucination*	0.484	*tribulation*	0.328	*hindering*	0.359

Table 1: Example entries for four (of the eight) emotions in the NRC Affect Intensity Lexicon. For each emotion, the table shows every 100th and 101th entry, when ordered by decreasing emotion intensity.

Linguistic Inquiry and Word Count

- Approximately 2300 words across 73 lexical resources associated with different psychological tasks
  - Pennebaker, J. W., Booth, R. J., and Francis, M. E. (2007). *Linguistic Inquiry and Word Count: LIWC 2007*. Austin,TX.
- Link:
  - https://www.liwc.app/
- Actively maintained and updated (most recent version is from 2022)
- Not free!

ANALYZE YOUR TEXT

Type or paste the text that you want to have analyzed into the box below. After you click "Analyze" you will receive a select set of LIWC-22 results for your text. There is currently a 5,000 character limit (approximately 1,000 words) on any given text. Your submitted text may be saved and used to fine-tune future versions of LIWC.

Note that this web demo is currently only able to analyze texts in the English language. The results that you receive from this online demo may differ slightly from the results calculated by the official LIWC-22 desktop application.

How would you classify this text? E-mail correspondence

Enter your text here:

Hi everyone,

Welcome to Week 11 of the semester! I hope you're all staying warm in the surprise Halloween snow. This week in class we're investigating additional ways to make sense of and extract information from text. Today we covered word sense disambiguation, and on Thursday we'll cover semantic role labeling. I just uploaded today's lecture recording to Blackboard.

ANALYZE TEXT

RESULTS

Traditional LIWC Dimension	Your Text	Average for E-mail Language
I-words (I, me, my)	2.10	1.85
Positive Tone	2.40	2.20
Negative Tone	0.00	0.50
Social Words	8.71	6.07
Cognitive Processes	10.81	11.40
Allure	4.50	4.83
Moralization	0.00	0.07
Summary Variables		
Analytic	71.65	75.99
Authentic	80.33	36.72

LIWC in Action

Brysbaert Concreteness Lexicon

- Approximately 40,000 words labeled with continuous concreteness labels ranging from 1-5
  - Brysbaert, M., Warriner, A.B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46, 904-911.
- Link:
  - http://crr.ugent.be/archives/1330

Word	Bigram	Conc.M	Conc.SD
roadsweeper	0	4.85	0.37
traindriver	0	4.54	0.71
tush	0	4.45	1.01
hairdress	0	3.93	1.28
pharmaceutics	0	3.77	1.41
hoover	0	3.76	1.23
shopkeeping	0	3.18	1.19
pushiness	0	2.48	1.24
underdevelop	0	2.37	1.4
tirelessness	0	2.28	1.28
oldfashioned	0	2.26	1.02
wellmannered	0	2.25	1.14
dismissiveness	0	1.83	1
spitefulness	0	1.8	0.76
untruthfulness	0	1.73	0.92
dispiritedness	0	1.56	0.71
sled	0	5	0
plunger	0	4.96	0.2
human	0	4.93	0.26
waterbed	0	4.93	0.27
cymbal	0	4.92	0.28
ginger	0	4.92	0.27
bobsled	0	4.9	0.41
cardboard	0	4.9	0.41
olive	0	4.9	0.31
dogsled	0	4.89	0.32

Personality and Stance

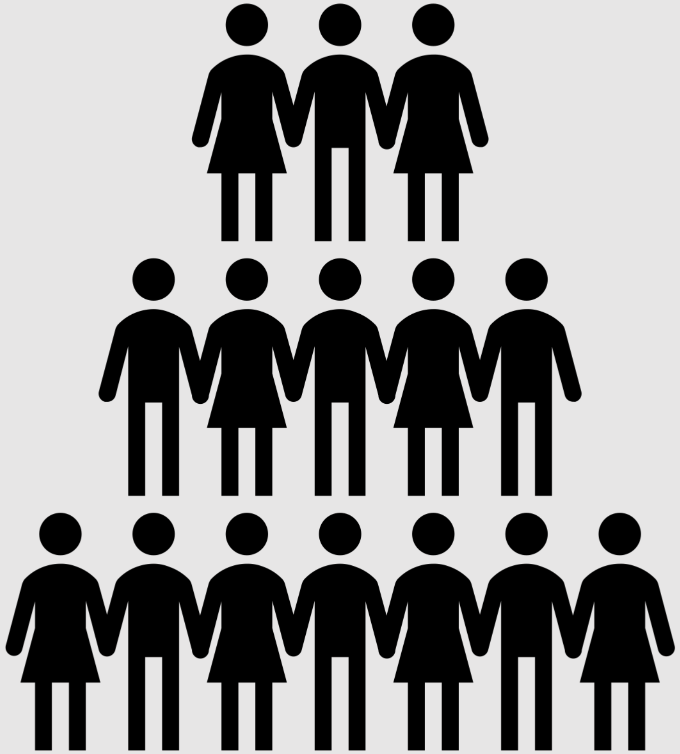
- Two other popular forms of affective analysis:
  - Personality detection
  - Stance detection
- Personality detection focuses on recognizing and classifying predefined aspects of a user's personal character
- Stance detection focuses on recognizing a user's opinion towards a specific topic

Personality Dimensions

- Most work in NLP makes use of the "Big Five" **personality dimensions**
  - Extroversion vs. Introversion
  - Emotional Stability vs. Neuroticism
  - Agreeableness vs. Disagreeableness
  - Conscientiousness vs. Unconscientiousness
  - Openness to Experience
- Several corpora exist containing language samples annotated with these personality dimensions
- Paper:
  - Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1), 417-440.

Affective Stance

- One's position towards another (or towards a topic) during an interaction
  - Friendly
  - Distant
  - Supportive
  - Unsupportive
- Corpora also exist containing conversational exchanges labeled with each party's affective stance



How to build a lexicon?

- Expert labels
  - Very reliable 🙂
  - Costly and time-consuming 😕
- Crowdsourced labels
  - Less reliable 😕
  - Inexpensive and quick 🙂

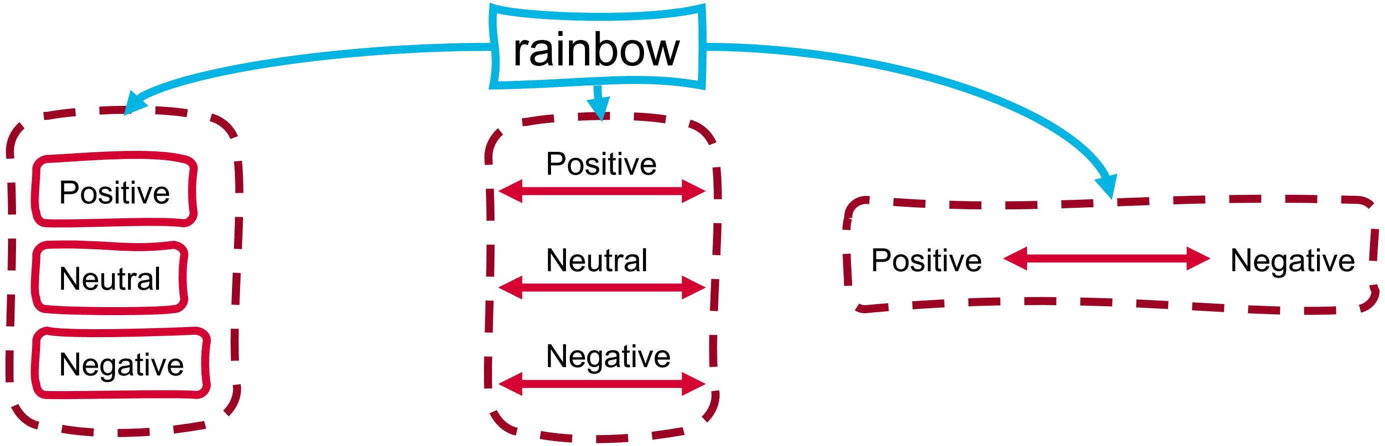
Crowdsourcing Resources

- Amazon Mechanical Turk:
  - https://www.mturk.com
- Appen:
  - https://appen.com
- Prolific:
  - https://www.prolific.com/
- Your own online survey

Annotation Schemata

- What kinds of labels will be permissible for your annotators?
- How will they know which labels to select?

Positive: A word that evokes a happy or content emotion.
Examples: *love*, *great*, *happy*
Neutral: A word that does not particularly evoke any emotion.
Examples: *pencil*, *refrigerator*, *khaki*
Negative: A word that evokes a sad or angry emotion.
Examples: *violence*, *evil*, *upset*



Adjudication

- Third-party adjudicator
- Majority label
- Average label

This Week's Topics

Tuesday

- Affective Analysis
- Affective Lexicons
- Inducing Affect and Association

Thursday

- Affective Tasks
- Coreference Resolution
- Referring Expressions

Semi-Supervised Induction of Affect Lexicons

- **Semi-supervised label induction:** The process of labeling new, unlabeled instances based on their similarity to instances in a small, labeled seed set
- Two main families:
  - **Axis-based** induction
  - **Graph-based** induction

Axis-Based Lexicon Induction

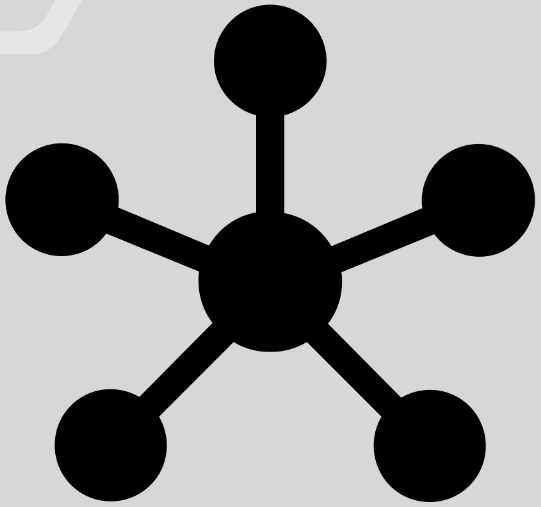
- Given a seed set, how similar is the instance to positive instances and how different is it from negative instances?
- The seed set may be:
  - Fine-tuned to the domain using induction techniques
  - Chosen manually to represent your domain

Axis-Based Lexicon Induction

- Once we've determined our seed words:
  - Compute an embedding for each seed word
  - Find the centroid of the embeddings for positive words, and the centroid of the embeddings for negative words
    - $\mathbf{V}^{+} = \frac{1}{n}\sum_{n}^{n} E(w_i^{+})$
  - Compute the axis by subtracting one centroid from another
    - $\mathbf{V}_{\boldsymbol{axis}} = \mathbf{V}^{+} - \mathbf{V}^{-}$
    - This produces a semantic axis vector that encodes the antonymy between the sets of words (in this case, creating a vector in the direction of the positive sentiment)
  - Compute the similarity between a given word embedding and the axis
    - $\text{score}(w) = \cos(E(w), \mathbf{V}_{axis}) = \frac{E(w) \cdot \mathbf{V}_{axis}}{\|E(w)\| \|\mathbf{V}_{axis}\|}$
  - Higher similarities indicate closer alignment with the positive class
- Paper: https://aclanthology.org/P18-1228.pdf

As an alternative….

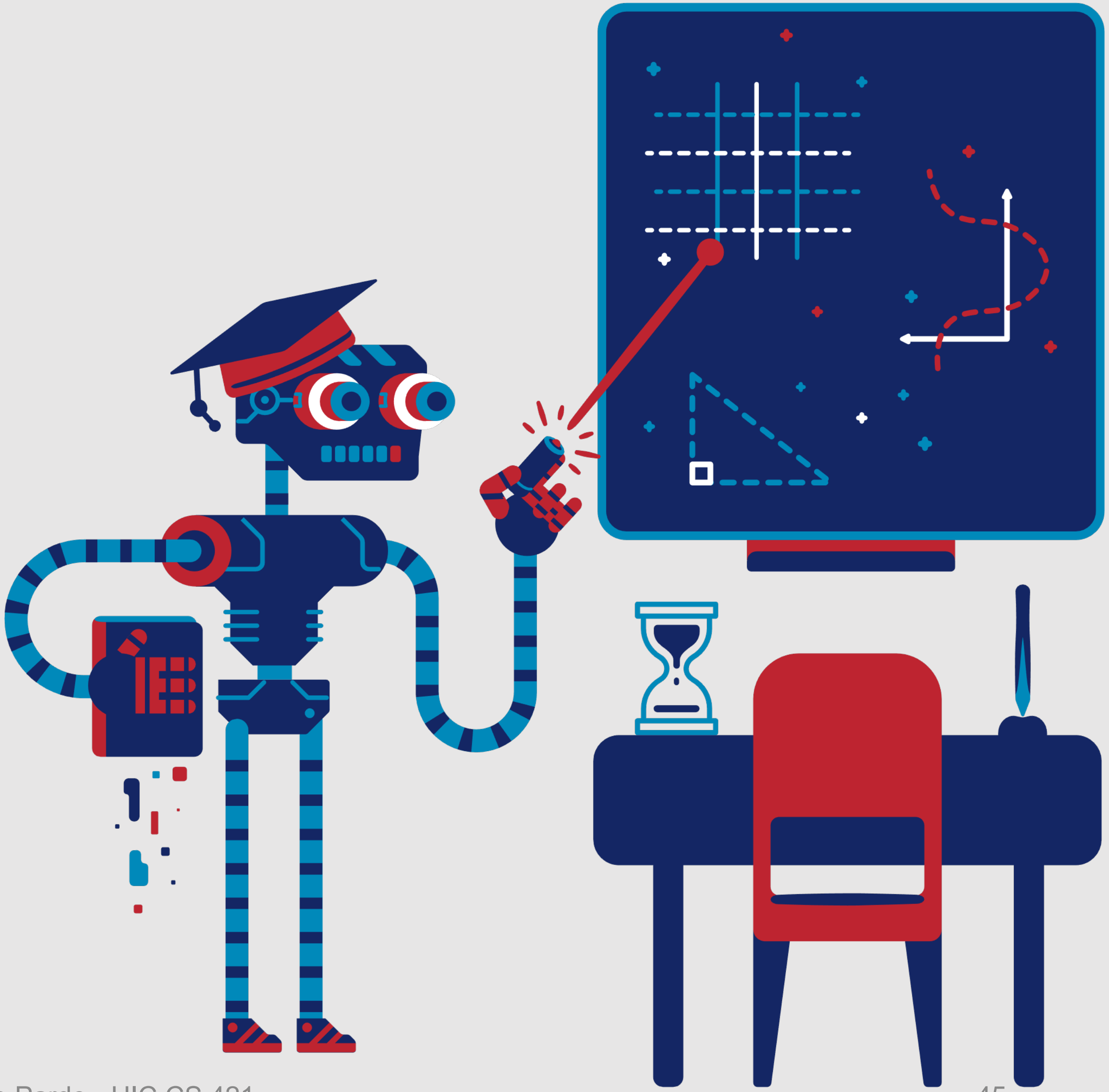
- Graph-based induction techniques allow us to define lexicons by propagating sentiment labels on graphs

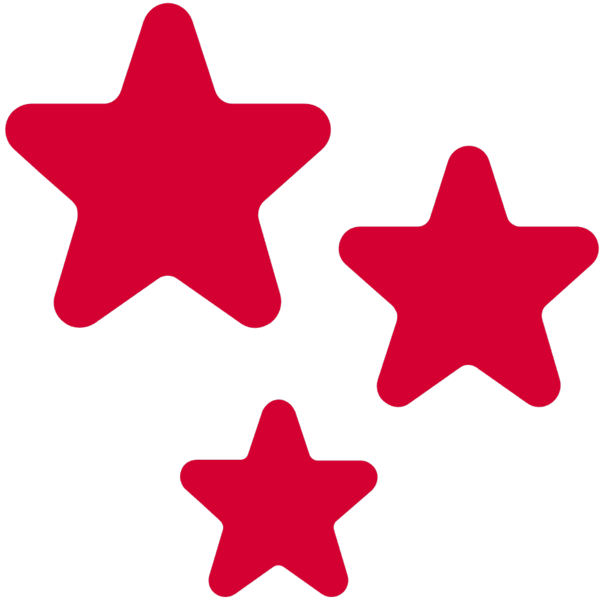


Graph-Based Lexicon Induction

- Given a graph that connects words with their nearest neighbors, how likely is it that a random walk from a positive word ends on the given word?
  - Define a graph that connects each word to its *k* nearest neighbors, with edges weighted by word similarity
  - Identify words in the graph belonging to a labeled seed set
  - Starting at a word from the seed set, perform an edge-weighted random walk
  - Assign an unlabeled word's score based on the probability of landing on it during a random walk from a positive seed and a random walk from a negative seed
    - $\text{score}^{+}(w_i) = \frac{\text{score}^{+}(w_i)}{\text{score}^{+}(w_i)+\text{score}^{-}(w_i)}$
  - Repeat multiple times using bootstrapping, and assign confidence to word scores based on their standard deviation across multiple runs

How can we use supervised machine learning to predict a word's sentiment?





Lots of supervision signals (potential labels) exist in real-world data.

- One example: Review scores
  - 1-5 stars
  - Rating from 1-10
  - Often associated with free-form review text
- We can use these scores and associated text to learn polarity distributions for words

Normalized Word Likelihood

- **Document-level sentiment classifier** → any statistical or neural methods we've learned about so far!
- **Word-level sentiment classifier** → also consider simple probabilistic measures
- **Normalized word likelihood**
  - $P(w|c) = \frac{\text{count}(w,c)}{\sum_{w \in C} \text{count}(w,c)}$

Normalized word likelihood helps us find a word's sentiment distribution across classes.

How likely the word is to be associated with one star, two stars, and so on!

We can then visualize this distribution using a **Potts diagram**

Potts Diagrams

- Mechanism for visualizing word sentiment
  - Sentiment class vs. normalized word likelihood
- Characteristic patterns:
  - **“J” shape:** Strongly positive word
  - **Reverse “J” shape:** Strongly negative word
  - **“Hump” shape:** Weakly positive or negative word
- Patterns may also correspond to different types of word classes
  - Emphatic and attenuating adverbs

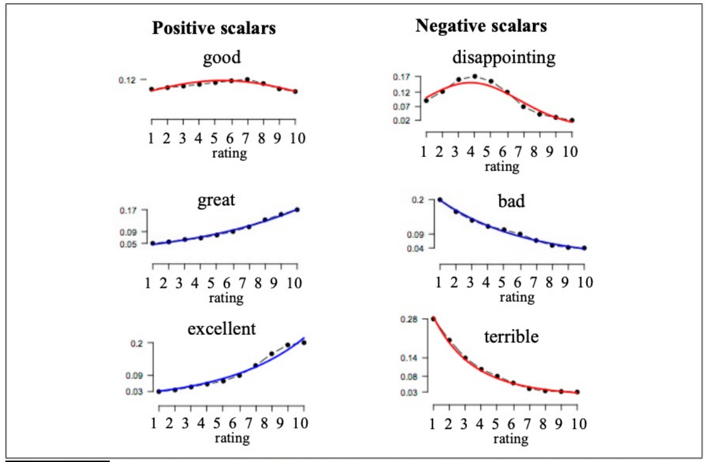


Figure 21.10 Potts diagrams (Potts, 2011) for positive and negative scalar adjectives, showing the J-shape and reverse J-shape for strongly positive and negative adjectives, and the hump-shape for more weakly polarized adjectives.

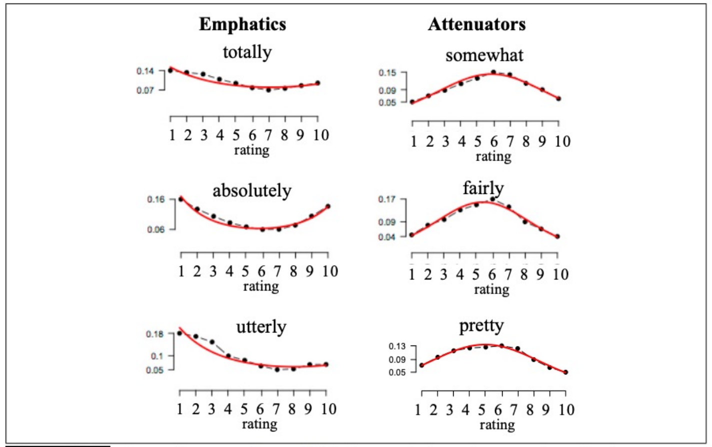


Figure 21.11 Potts diagrams (Potts, 2011) for emphatic and attenuating adverbs.

Sometimes we want to figure out which words are most closely associated with a class rather than just plotting frequency distributions.

- Certain words will have high frequencies across many classes, but these are less helpful for understanding class membership
- Certain words will have low frequencies across all classes, but still might be more informative if they are more clearly associated with a specific class
- More sophisticated approach: **Log odds ratio with an informative Dirichlet prior**

Log Odds Ratio

- Originally proposed for measuring partisan speech used by US politicians
  - Some words are likelier to be used by Republicans, and other words are likelier to be used by Democrats
- Generalizes to any other problem domain for which lexical trends are anticipated to be different
- Key goal: Find words that are statistically overrepresented in one category of text compared to another

Start with a simple log odds ratio….

- **Probability of word *w* existing in a subset of words *i*:**
  - $P^i(w) = \frac{f_w^i}{n^i}$

Total number of words in *i*

- **Log odds ratio for word *w* in the subset of words *i* versus the subset of words *j*:**
  - $\text{lor}(w) = \log \frac{P^i(w)}{1-P^i(w)} - \log \frac{P^j(w)}{1-P^j(w)} = \log \frac{f_w^i}{n^i - f_w^i} - \log \frac{f_w^j}{n^j - f_w^j}$

- However, this doesn't tell us anything about what we'd expect to see!

Dirichlet Intuition

- Use a large background corpus to get a prior estimate of our expected frequency for each word *w*
- To do so:
  - Add the counts from that corpus to our numerator and denominator
  - This basically shrinks the counts toward that prior (how big are the differences, *given what we would expect from a large background corpus*?)

Prior-Modified Log Odds Ratio

- Modifying the previous equation with an informative Dirichlet prior:
  - $\delta_w^{(i-j)} = \log \frac{f_w^i + \alpha_w}{n^i + \alpha_0 - (f_w^i + \alpha_w)} - \log \frac{f_w^j + \alpha_w}{n^j + \alpha_0 - (f_w^j + \alpha_w)}$

Number of words in *i*

Count of *w* in corpus

Size of background corpus

Count of w in background corpus

Log Odds Ratio with Informative Dirichlet Prior

- Estimate of variance for the modified log odds ratio:
  - $\sigma^2\left(\hat{\delta}_w^{(i-j)}\right) \approx \frac{1}{f_w^i+\alpha_w} + \frac{1}{f_w^j+\alpha_w}$
- Final statistic for a word is then the z-score of its modified log odds ratio:
  - $\frac{\hat{\delta}_w^{(i-j)}}{\sqrt{\sigma^2\left(\hat{\delta}_w^{(i-j)}\right)}}$

Key Modifications to Log Odds Ratio

- This method thus modifies the standard version of log odds ratio by:
  - Using z-scores of the log odds ratio
    - Controls for the amount of variance in a word's frequency
  - Using counts from a background corpus to provide a prior count for words
    - Provides expected differences to contextualize the observed differences

This ultimately gives us a useful tool for analysis!

Words associated with one-star restaurant reviews

worst, rude, terrible, horrible, bad, awful, disgusting, bland, tasteless, gross, mediocre, overpriced, worse, poor

Great, best, love, delicious, amazing, favorite, perfect, excellent , awesome, friendly, fantastic, fresh, wonderful, incredible, sweet, yum

Words associated with five-star restaurant reviews

Summary: Affective Lexicons

- Lexicons can help us distinguish many kinds of **affective states**
- Emotion can be represented using fixed **atomic units** or **dimensions** in a continuous space
- **Affective lexicons** can be built by hand, in a semi-supervised manner, or using fully supervised methods
- Words can be assigned weights in a lexicon based on frequency measures and ratio metrics like **log odds ratio with an informative Dirichlet prior**

This Week's Topics

Affective Analysis
Affective Lexicons
Inducing Affect and Association

Tuesday

Thursday

Affective Tasks
Coreference Resolution
Referring Expressions

What is affect recognition?

- Affect recognition: The task of automatically determining how a given input makes would be characterized, based on some specified range of categories
  - Happy vs. sad
  - Extroverted vs. introverted
  - Friendly vs. distant

Affect Recognition

- Typically framed as a supervised learning task
- Smaller datasets:
  - N-gram features
- Larger datasets:
  - N-gram features, pruned based on frequency or **pointwise mutual information (PMI)**
    - $\mathrm{PMI}(x;y) = \log \frac{p(x,y)}{p(x)p(y)}$
- Even larger datasets:
  - Word embeddings

Features from External Lexicons

- **Indicator Function:**
  - $f_L(x) = \begin{cases} 1 & if\ \exists w : w \in L \,\&\, w \in x \\ 0 & \text{otherwise} \end{cases}$
- **Count-Based Function:**
  - $f_L(x) = \sum_{w \in x} \text{count}_L(w)$

Can vary depending on the lexicon (e.g., a binary increment or a continuous score)

Lexicon-based features can shed new light on interesting social science problems!

- Does one’s use of positive language correlate with one’s level of extroversion?
- Is more concrete language likely to evoke more neutral emotions?
- Is there a relationship between the number of “difficult” words and the overall subjectivity of an input?

What if we don’t have labeled training data to build a supervised model for sentiment or affect recognition?

Use sentiment resources (such as those already described!) to perform sentiment analysis directly

Using Lexicons for Sentiment Recognition

1

Assign a positive label to instances that contain more positive than negative words from the lexicon

2

Assign a negative label to instances that contain more negative than positive words

3

Assign a neutral label in the event of a tie

More formally....

- Define a threshold λ indicating the minimum percentage of positive or negative words needed for a positive or negative classification
- Select a sentiment class as follows:
  - $f^+ = \sum_{w \in x} \theta_w^+ \text{count}_+(w)$
  - $f^- = \sum_{w \in x} \theta_w^- \text{count}_-(w)$
  - $\text{sentiment} = \begin{cases} + \text{ if } \frac{f^+}{f^-} > \lambda \\ - \text{ if } \frac{f^-}{f^+} > \lambda \\ 0 \text{ otherwise} \end{cases}$

Entity-Centric Affect

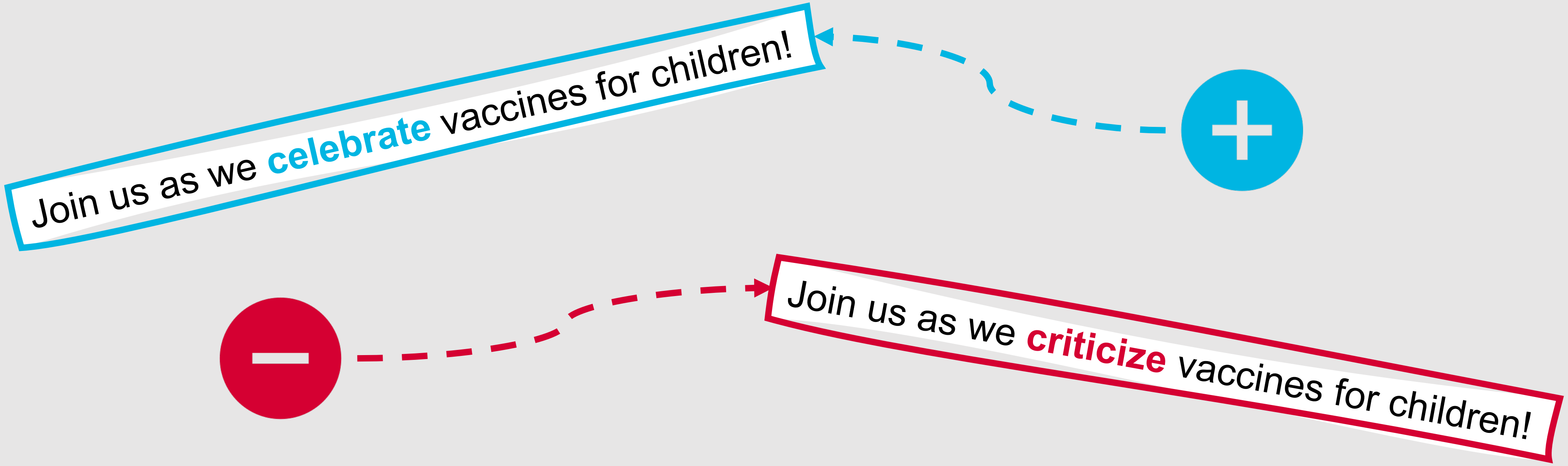
- Sometimes we don't need (or want) to recognize the affect of an entire input
  - Scope may be too broad!
- We can also learn to predict the affect of a single entity within the input

Methods for Entity-Centric Affect Recognition

One way to do this: Leverage both affect lexica and contextual word embeddings	First, extract a contextual embedding for each instance of a word
	Then, average those embeddings
	Repeat this process for all words
	Learn to map from averaged word embeddings to affective scores
When a new entity is encountered without a known affective score:	Create a new average embedding based on all instances of that entity in context
	Predict a score for that embedding

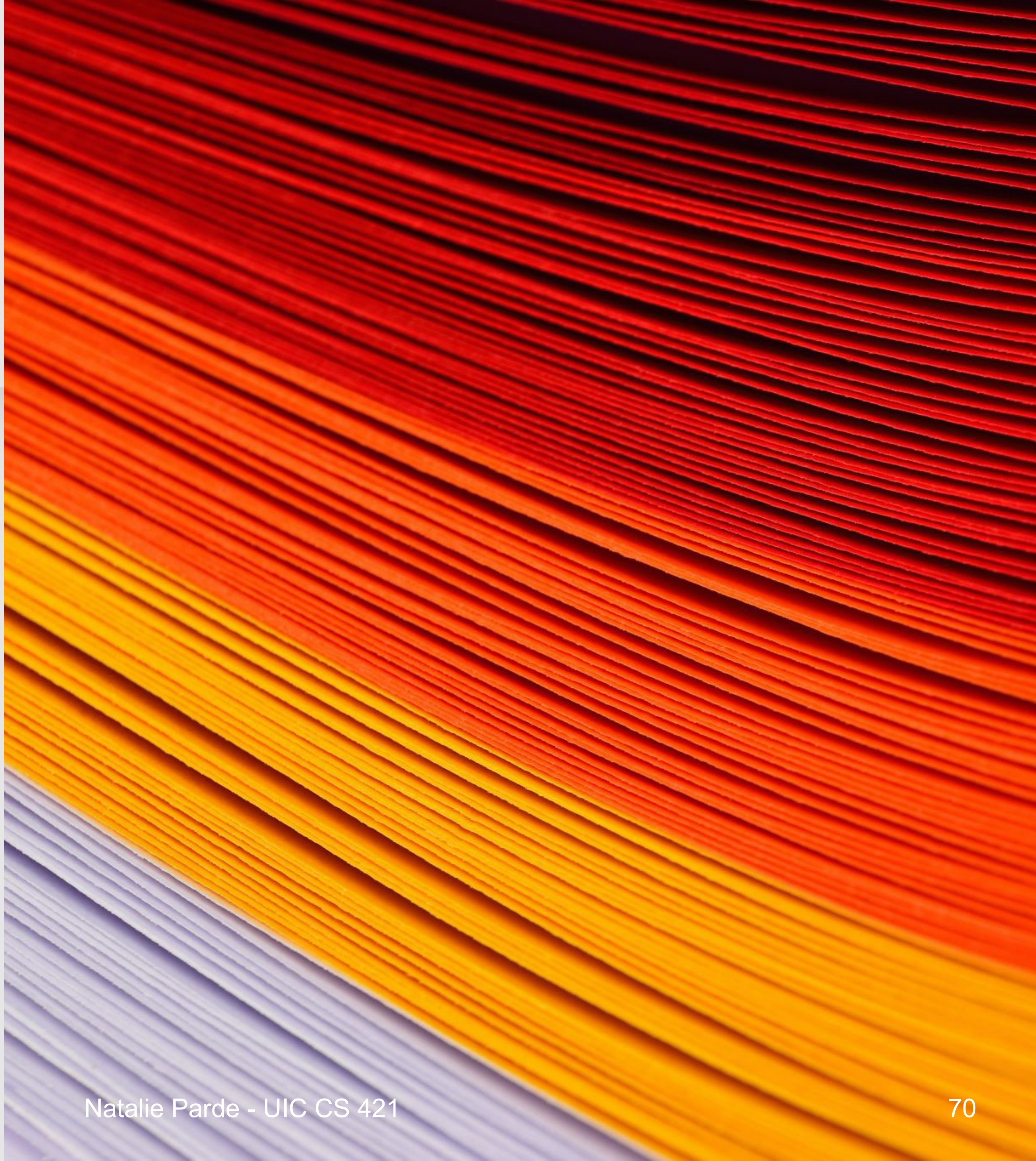
Connotation Frames

- We can also represent affective meaning using a multidimensional affective space *or* using **connotation frames**
  - Indicate affective properties commonly associated with words, similarly to how verb frames indicate selectional preferences



Connotation Frames

- Formalism for analyzing subjective roles and relationships implied by a predicate
- Contain labels for relationships that re inferable from the predicate:
  - Writer’s perspective
  - Reader’s perspective
  - Entity’s perspective
  - Entity’s value
  - Entity’s mental state
  - Effect on entity

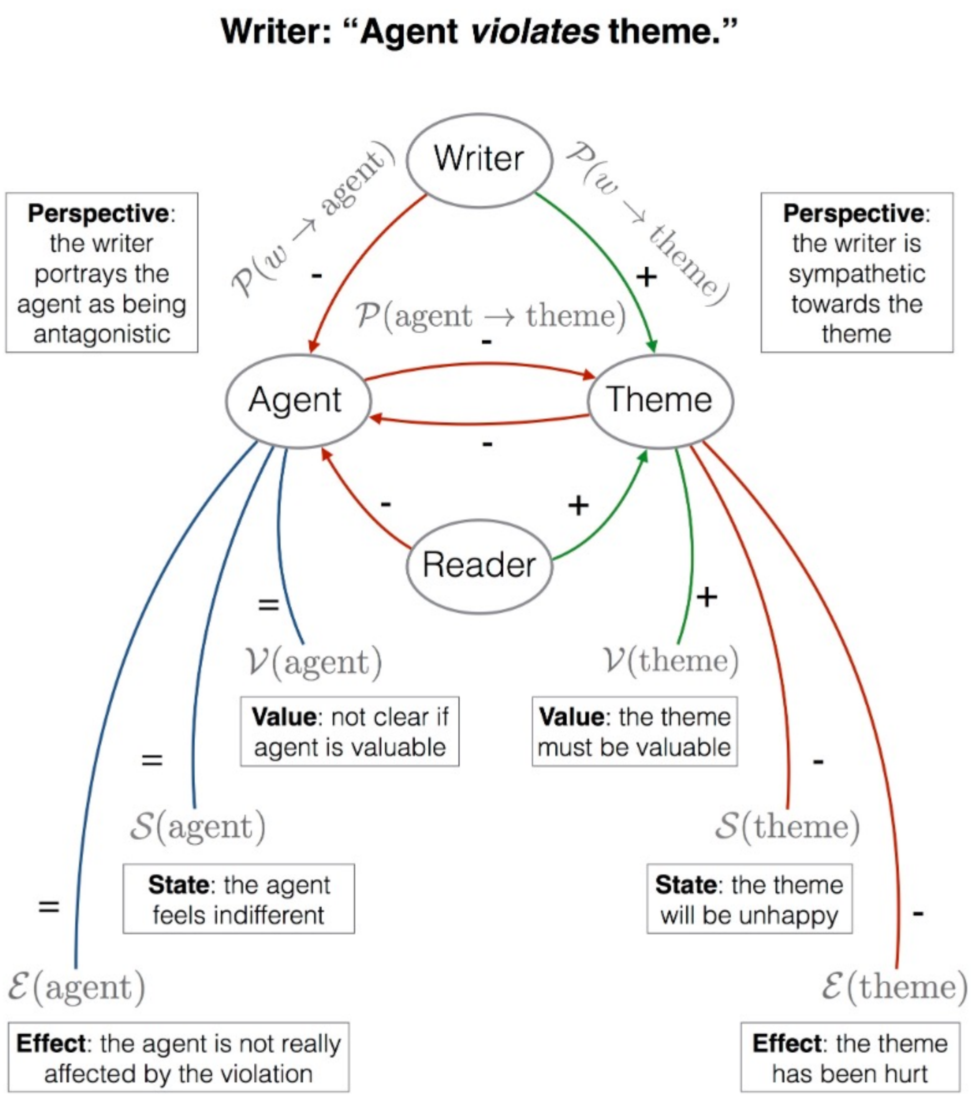


Connotation Frames

- Can be:
  - Manually constructed
  - Learned automatically
- Downloadable collection of connotation frames:
  - https://hrashkin.github.io/connframe.html

Example Connotation Frame

- Relevant papers:
  - Hannah Rashkin, Sameer Singh, Yejin Choi. 2016. Connotation Frames: A Data-Driven Investigation. In Proceedings of ACL 2016.
    - https://aclanthology.org/P16-1030/
  - Maarten Sap, Marcella Cindy Prasettio, Ari Holtzman, Hannah Rashkin, & Yejin Choi. 2017. Connotation Frames of Power and Agency in Modern Films. In Proceedings of EMNLP 2017 Short Papers.
    - https://aclanthology.org/D17-1247/
- Relevant resource for measuring social dynamics between personas: https://github.com/maartensap/riveter-nlp



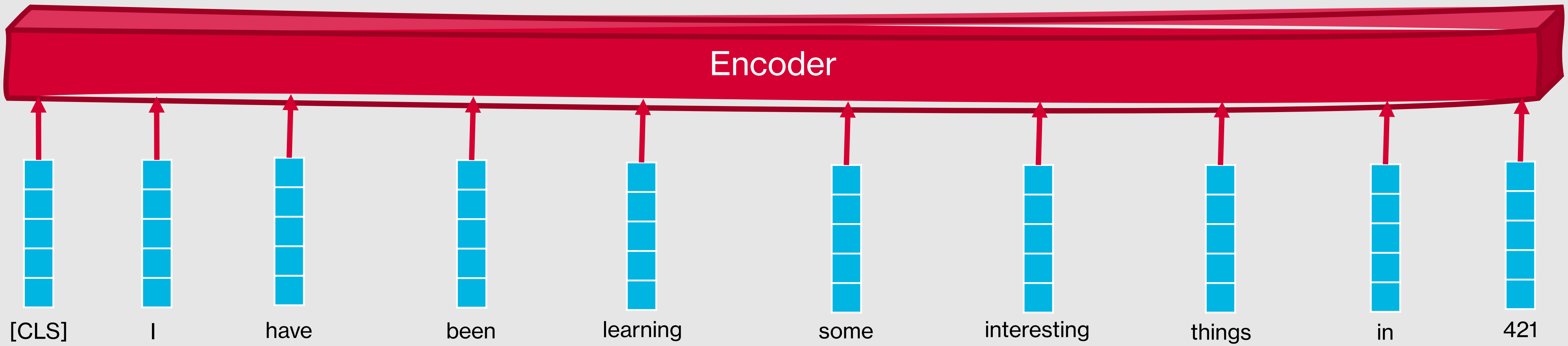
Stance Detection

- The task of determining sentiment with respect to a specific target (often a polarizing topic)
- Stance labels are generally some variation of **favor** and **against**
- Numerous datasets exist for this task:
  - **SemEval-2016 Task 6A Stance Dataset**
    - 4870 tweets manually annotated for stance with respect to: “Atheism,” “Climate Change is Real Concern,” “Feminist Movement,” “Hilary Clinton,” and “Legalization of Abortion”
  - **Multi-Perspective Consumer Health Query Data**
    - Relevant sentences from the top 50 links corresponding to common, polarizing or widely debated public health queries (e.g., “Does the MMR vaccine lead to autism in children?”)
  - **Ideological Online Debates**
    - Online debates on “Existence of God,” “Healthcare,” “Gun Right,” “Gay Rights,” and “Abortion and Creationism”

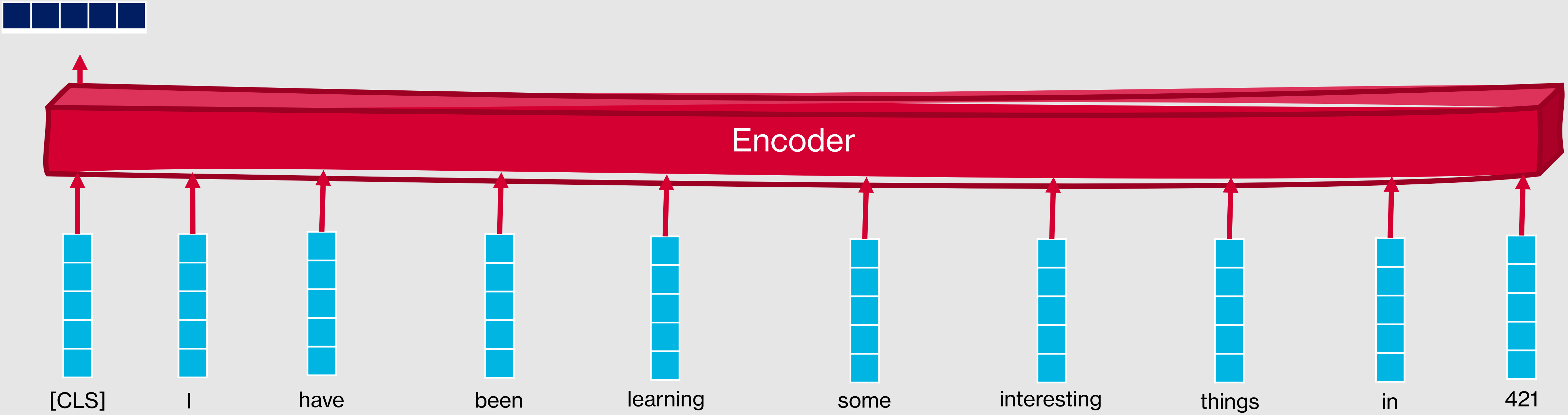
How can we build stance detection models?

- Predict favor and against (and optionally neutral) labels for each target *or* determine the relevant target as a preliminary step
- Feature-based approaches:
  - Text content
  - (If known) user-specific attributes
  - (If known) network-specific attributes
- Neural approaches:
  - Typically framed as a supervised instance-level classification task

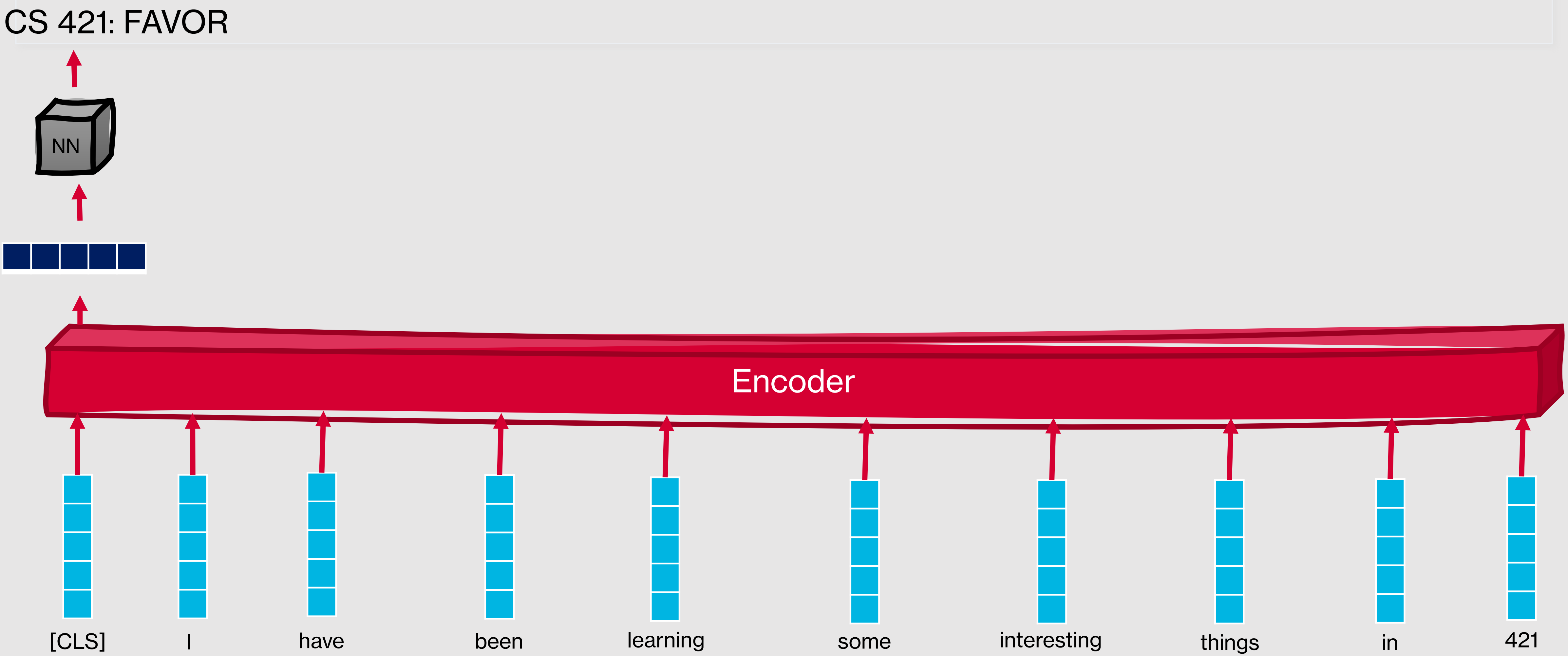
Example: Stance Detection



Example: Stance Detection



Example: Stance Detection



This Week's Topics

Tuesday

- Affective Analysis
- Affective Lexicons
- Inducing Affect and Association

Thursday

- Affective Tasks
- Coreference Resolution
- Referring Expressions

What is coreference resolution?

The process of automatically identifying expressions that refer to the same entity

Coreference resolution is essential to creating high-performing NLP systems.

Which NLP course do you want to take next year?

What are my options?

Well, there's CS 421: Natural Language Processing, CS 521: Statistical Natural Language Processing, CS 532: Advanced Topics in NLP, and CS 533: Deep Learning for NLP.

Hmm, I'll do Statistical NLP.

Coreference resolution is essential to creating high-performing NLP systems.

Which NLP course do you want to take next year?

What are my options?

Well, there's CS 421: Natural Language Processing, **CS 521: Statistical Natural Language Processing**, CS 532: Advanced Topics in NLP, and CS 533: Deep Learning for NLP.

Hmm, I'll do **Statistical NLP**.

Both humans and NLP systems interpret language with respect to a discourse model.

- **Discourse model:** Mental model that is built incrementally, containing representations of entities, their properties, and the relations between them
- **Referent:** The discourse entity itself
  - (CS 521: Statistical Natural Language Processing)
- **Referring expression:** The linguistic expression referring to a referent
  - "CS 521"
  - "CS 521: Statistical Natural Language Processing"
  - "521"
  - "Statistical NLP"
- Two or more referring expressions that refer to the same discourse entity are said to **corefer**

Anaphora

- **Anaphora:** Referring to an entity that has already been introduced in the discourse
  - First mention is the **antecedent**
  - Subsequent mentions are **anaphors**
  - Entities with only a single mention are **singletons**

The University of Illinois at Chicago is an excellent place to study natural language processing. UIC has many faculty currently working in the area, including but not limited to Natalie Parde, Barbara Di Eugenio, Cornelia Caragea, Bing Liu, and Philip Yu. The school is located in bustling downtown Chicago, and as a bonus it will be opening a snazzy new CS building in 2024.

Anaphora

- **Anaphora:** Referring to an entity that has already been introduced in the discourse
  - First mention is the **antecedent**
  - Subsequent mentions are **anaphors**
  - Entities with only a single mention are **singletons**

The **University of Illinois at Chicago** is an excellent place to study natural language processing. UIC has many faculty currently working in the area, including but not limited to Natalie Parde, Barbara Di Eugenio, Cornelia Caragea, Bing Liu, and Philip Yu. The school is located in bustling downtown Chicago, and as a bonus it will be opening a snazzy new CS building in 2024.

Anaphora

- **Anaphora:** Referring to an entity that has already been introduced in the discourse
  - First mention is the **antecedent**
  - Subsequent mentions are **anaphors**
  - Entities with only a single mention are **singletons**

The **University of Illinois at Chicago** is an excellent place to study natural language processing. **UIC** has many faculty currently working in the area, including but not limited to Natalie Parde, Barbara Di Eugenio, Cornelia Caragea, Bing Liu, and Philip Yu. **The school** is located in bustling downtown Chicago, and as a bonus **it** will be opening a snazzy new CS building in 2024.

Anaphora

- **Anaphora:** Referring to an entity that has already been introduced in the discourse
  - First mention is the **antecedent**
  - Subsequent mentions are **anaphors**
  - Entities with only a single mention are **singletons**

The **University of Illinois at Chicago** is an excellent place to study natural language processing. **UIC** has many faculty currently working in the area, including but not limited to **Natalie Parde**, Barbara Di Eugenio, Cornelia Caragea, Bing Liu, and Philip Yu. **The school** is located in bustling downtown Chicago, and as a bonus **it** will be opening a snazzy new CS building in 2024.

Coreference Chains

A set of coreferring expressions is often called a **coreference chain**

The **University of Illinois at Chicago** is an excellent place to study natural language processing. **UIC** has many faculty currently working in the area, including but not limited to **Natalie Parde**, Barbara Di Eugenio, Cornelia Caragea, Bing Liu, and Philip Yu. **The school** is located in bustling downtown Chicago, and as a bonus **it** will be opening a snazzy new CS building in 2024.

{"University of Illinois at Chicago", "UIC", "The school", "it"}

{"Natalie Parde"}

Two Key Tasks

- **Coreference resolution** thus generally comprises two key tasks:
  - Identify **referring expressions** (mentions of entities)
  - Cluster them into **coreference chains**
- We can also perform **entity linking** to map coreference chains to real-world entities
  - {“University of Illinois at Chicago”, “UIC”, “The school”, “it”} → https://en.wikipedia.org/wiki/University_of_Illinois_at_Chicago

This Week's Topics

Tuesday

- Affective Analysis
- Affective Lexicons
- Inducing Affect and Association

Thursday

- Affective Tasks
- Coreference Resolution
- Referring Expressions

Linguistic Background

- Referring expressions can occur in several forms:
  - **Indefinite noun phrases**
  - **Definite noun phrases**
  - **Pronouns**
  - **Proper nouns (names)**
- These can be used to **evoke** and **access** entities in the discourse model in a variety of ways

Indefinite Noun Phrases

- Usually marked with the determiner *a* or *an*
- Can also be marked with other indefinite terms
  - E.g., *some*
- Generally introduce **new entities** to the discourse

The blue line was experiencing delays so I took **an** Uber.

Definite Noun Phrases

- Usually marked with *the*
- Generally refer to entities that have already been introduced to the discourse
- May refer to entities that haven't been introduced to the discourse, but are identifiable to the receiver due to:
  - World knowledge
  - Implications from the discourse structure

The blue line was experiencing delays so I took **an** Uber. Unfortunately, so did everyone else …**the** Uber got stuck in a traffic jam.

Have you checked out **the** Andy Warhol exhibit?

Make sure to order **the** tiramisu!

Pronouns

- Generally refer to entities that have already been introduced to the discourse and are easily identifiable

> The blue line was experiencing delays so I took **an** Uber. Unfortunately, so did everyone else …**the** Uber got stuck in a traffic jam. **It** ended up reaching UIC later than the original train I'd been hoping to catch.

Proper Nouns (Names)

- Can be used either to introduce new entities to the discourse, or to refer to those that already exist

Chicago is one of the largest cities in the United States. **Chicago** is known for its architecture, its thriving arts and music scene, its hot dogs and deep dish pizza, and---of course---its winter weather.

Information Status

- Referring expressions can also be categorized by their **information status**
  - The way they introduce **new information** or access **old information**
- Three main groups:
  - New noun phrases
  - Old noun phrases
  - Inferables

New Noun Phrases

- **Brand new NPs:** Introduce entities that are both **new to the discourse** and **new to the listener**
  - E.g., *an Uber*
- **Unused NPs:** Introduce entities that are **new to the discourse** but **not to the listener**
  - E.g., *Chicago*

Old Noun Phrases

- Introduce entities that already exist in the discourse model (and are thus **not new to the discourse nor to the listener**)
  - E.g., *she*

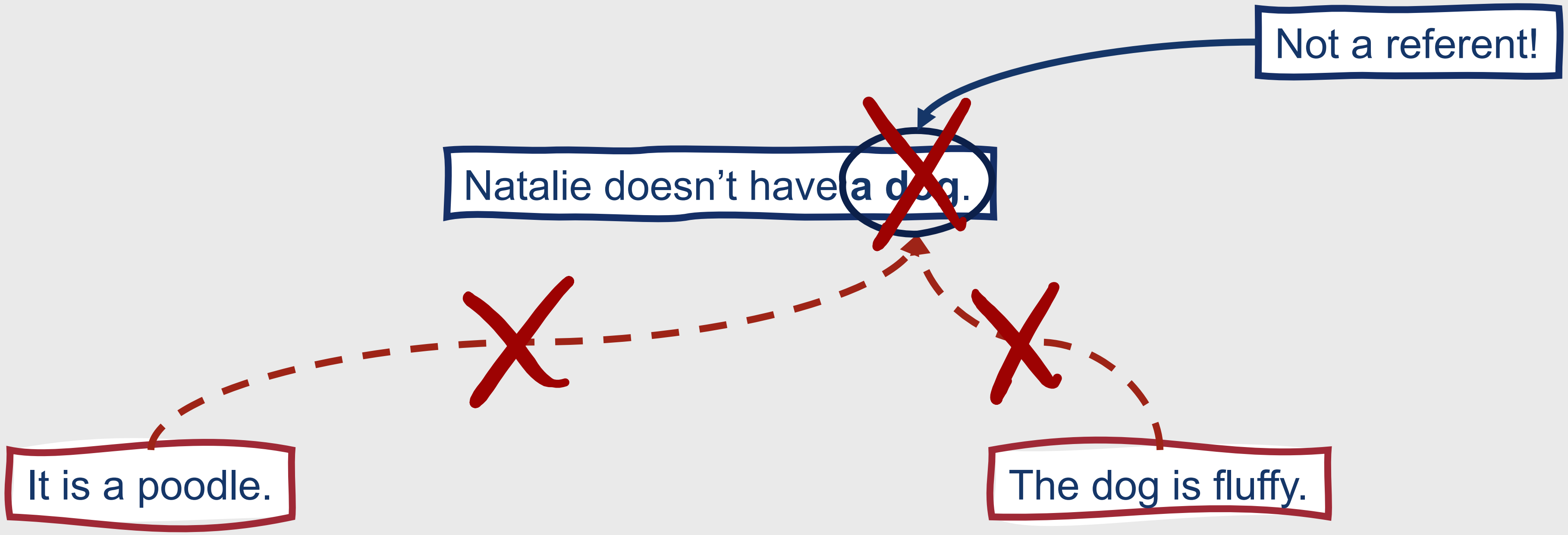
Inferables

- Introduce entities that are **new to the discourse** and **new to the listener** *but* the hearer can infer their existence by reasoning about other entities already introduced
  - E.g., I got in my Uber and told *the driver* to take us to UIC as fast as she could.

Generally, the form of a referring expression gives strong clues about its information status.

- **Very salient** (easily accessible) entities can be referred to using **less linguistic material**
  - E.g., pronouns
- **Less-salient** entities (e.g., those that are discourse-new and hearer-new) require **more linguistic material**
  - E.g., full names

Note: Not all noun phrases are referring expressions!



Structures Easily Confused with Referring Expressions

Appositives	Noun phrases that describe other noun phrases	Natalie Parde, *Assistant Professor of Computer Science*, teaches CS 421.
Predicative and Prenominal Noun Phrases	Noun phrases that describe characteristics of other noun phrases	Natalie Parde is an *Assistant Professor*.
Expletives	Non-referential pronouns	Natalie thought *it* was cool that so many students at UIC were interested in NLP.
Generics	Pronouns that refer to classes of nouns in general, rather than specific instances of those nouns	In Chicago, *you* get to experience all four seasons - summer, early winter, winter, and late winter.

So far, we've focused on linguistic properties of referring expressions….

- What about linguistic properties of coreference relations (relations between an anaphor and its antecedent)?
    - Number agreement
    - Person agreement
    - Gender/noun class agreement
    - Binding theory constraints
    - Recency
    - Grammatical role
    - Verb semantics
    - Selectional restrictions

Number Agreement

- In general, antecedents and their anaphors should agree in number
  - Singular with singular
  - Plural with plural
- A few exceptions:
  - Some semantically plural entities (e.g., companies) can be referred to using either singular or plural pronouns
  - “They” can be used as a singular pronoun

Person Agreement

In general, antecedents and their anaphors should agree in person

First person with first person

- I, my, me

Third person with third person

- They, their, them

An exception:

Text containing quotations

- "**I** spent twelve hours making those slides," **she** pointed out.

Gender/Noun Class Agreement

- In general, antecedents and their anaphors should agree in grammatical gender
  - He with his
  - She with hers
  - They with theirs
- This is an even bigger deal in languages for which all nouns have grammatical gender
  - La casa 🏠
  - El banco 🏦

Binding Theory Constraints and Recency

- **Binding Theory Constraints:** Antecedents and their anaphors should adhere to the syntactic constraints placed upon them
  - Reflexive pronouns (e.g., herself) corefer with the subject of the most immediate clause that contains them
    - **Natalie** told **herself** that she wouldn't be nearly as busy next week.
- **Recency:** Antecedents introduced recently tend to be more salient than those introduced earlier
  - Pronouns are likelier to be anaphors for the most recent plausible antecedent
    - Natalie went to a **faculty meeting**. Shahla went to a **student government meeting**. **It** was mainly about new policy changes that had recently been approved.

Grammatical Role

- Antecedents in some grammatical roles are more salient than others
  - Subject position > object position
    - **Natalie** went to the Eiffel Tower with **Shahla**. **She** took a selfie.

Verb Semantics

- Salience may be influenced by the types of verbs to which antecedents and anaphors are arguments
  - **Natalie** congratulated **Shahla**. **Her** paper had just been accepted.
  - **Natalie** bragged to **Shahla**. **Her** paper had just been accepted.

Selectional Restrictions

- Finally, salience may also be influenced by other semantic knowledge about the verbs to which antecedents and anaphors are arguments
  - Natalie pulled her **suitcase** out of the **Uber**. **It** sped off into the sunset.

Summary: Affective Tasks and Referring Expressions

- **Connotation frames** express richer affective relationships, similar to those seen with semantic frames
- **Stance detection** allows us to predict sentiment with respect to a specific target
- **Coreference resolution** is the process of automatically identifying expressions that refer to the same entity
- This involves two tasks:
  - Identifying **referring expressions**
  - Clustering them into **coreference chains**