

# Word Embeddings (II)

CS 485, Fall 2026

Applications of Natural Language Processing

Brendan O'Connor

College of Information and Computer Sciences

University of Massachusetts Amherst

*[With slides from SLP3, Mohit Iyyer, and Laure Thompson]*

# Announcements

- HW1 due tonight: good luck finishing!
- Canvas "Discussions" board should be open - feel free to ask questions!
- Quick HW1 questions?
  - PDF export: "jupyter nbconvert", Jupyter/Colab "File->Export". Or convert to html then pdf. Or paste screenshots into a word processor, whatever is readable.

# Announcements

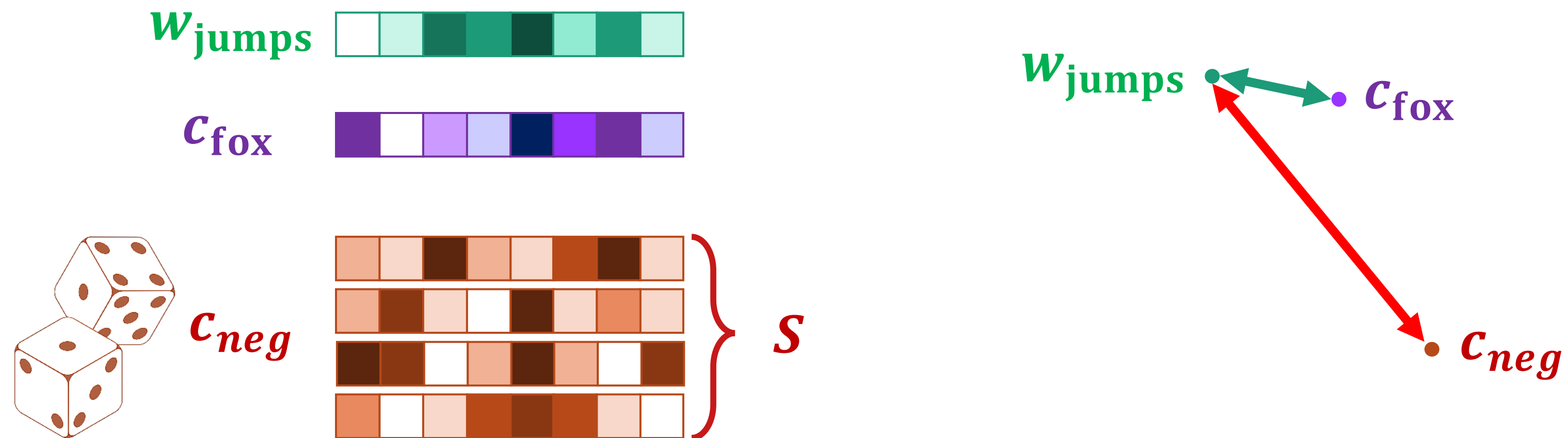
- HW2!
  - Releasing tomorrow, due next week, **Friday 10/9**. Intended to be a shorter homework!
  - Mostly math-y questions: Logistic regression, precision/recall, word embeddings and cosines
  - Also simple experiment to explore pretrained word embeddings
- Next week I'll be traveling to two NLP conferences. Changes:
  - Mon. 10/5 2:30pm, in this room: Nguyen will hold office hours, and review any material you all have questions about
  - Wed. 10/7 2:30pm, on Zoom: Shufan will hold office hours.
  - Fri. 10/9, 12:30pm to 1:30pm, hybrid: Nguyen will hold an additional office hours (it's the HW2 due date)
  - All 3 of us will not hold the "normal" office hours times



- We have
  - Dense vector model of word type representations (word "meanings")
  - For many words, learned from a large corpus
  - Learned from principle of distributional similarity
- What do distributional word embeddings learn?
- What can you do with them?

# SGNS: Summarizes target word's contexts

Co-occurrence **jumps** , **fox**:



- During training....
  - Move  $U_w$  closer to its context term vectors
$$P(+ | (w,c)) = g(\mathbf{U}_w' \mathbf{V}_c)$$
  - Move  $U_w$  away from non-occurring context term vectors
$$P(- | (w,c)) = 1 - g(\mathbf{U}_w' \mathbf{V}_c) = g(-\mathbf{U}_w' \mathbf{V}_c)$$
  - (Context term vectors  $V_c$  get updated as well)

- Demo! Let's look at GLoVE (similar to word2vec) embeddings
  - fasttext, glove, and word2vec are some often used pretrained word embeddings

# Linear semantic properties?

- Hypothesis: semantic properties correspond to directions in the word space
- Intuition: a property may correspond to certain contexts being more probable
- Toy example: male-indicating context words may corefer with [TARGET], indicating gender
  - [TARGET] . Then he ...
  - [TARGET] , a man ...

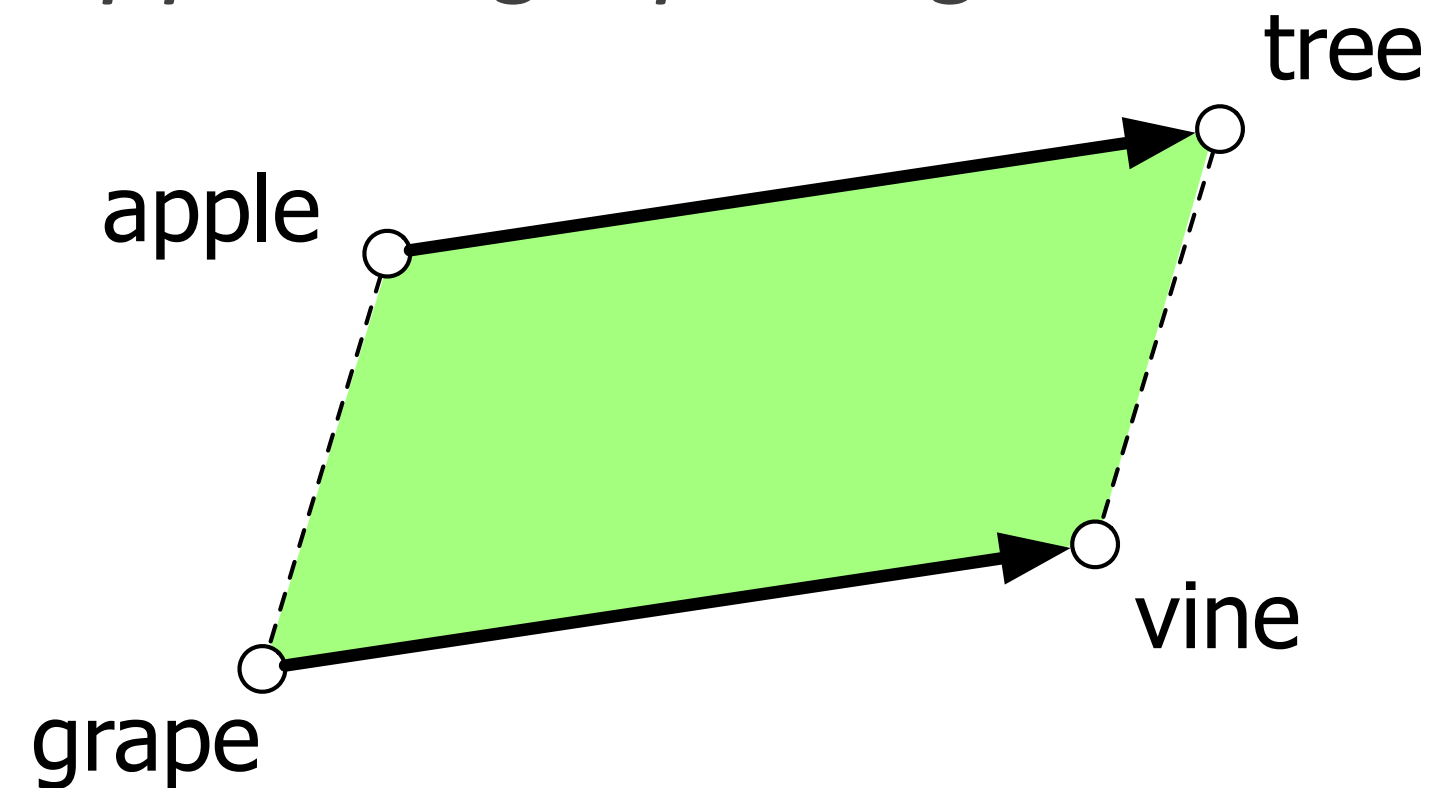
# Linear semantic properties?

## Analogical relations

The classic parallelogram model of analogical reasoning  
(Rumelhart and Abrahamson 1973)

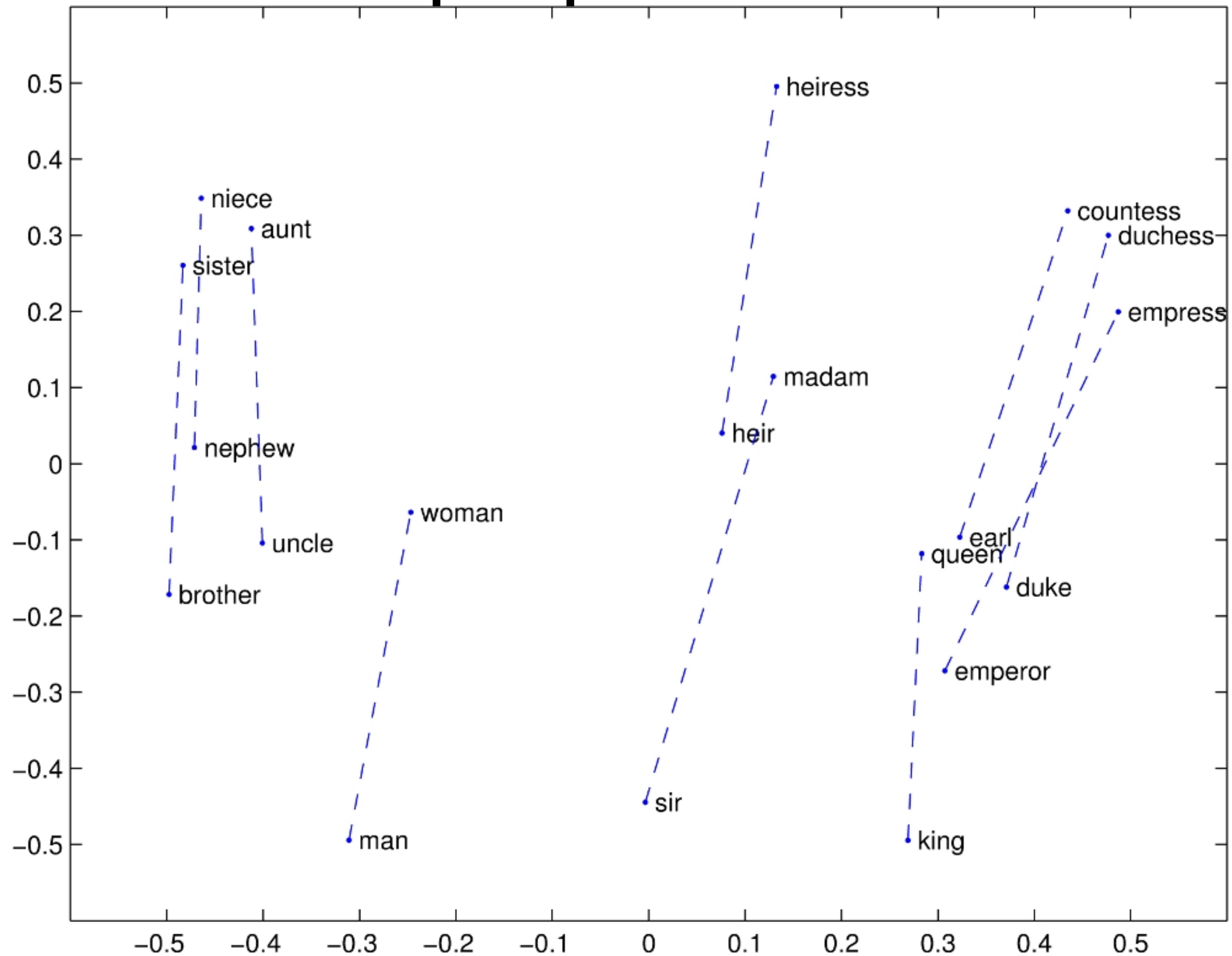
To solve: *"apple is to tree as grape is to \_\_\_\_\_"*

Add  $\overrightarrow{tree - apple}$  to  $\overrightarrow{grape}$  to get ***vine***



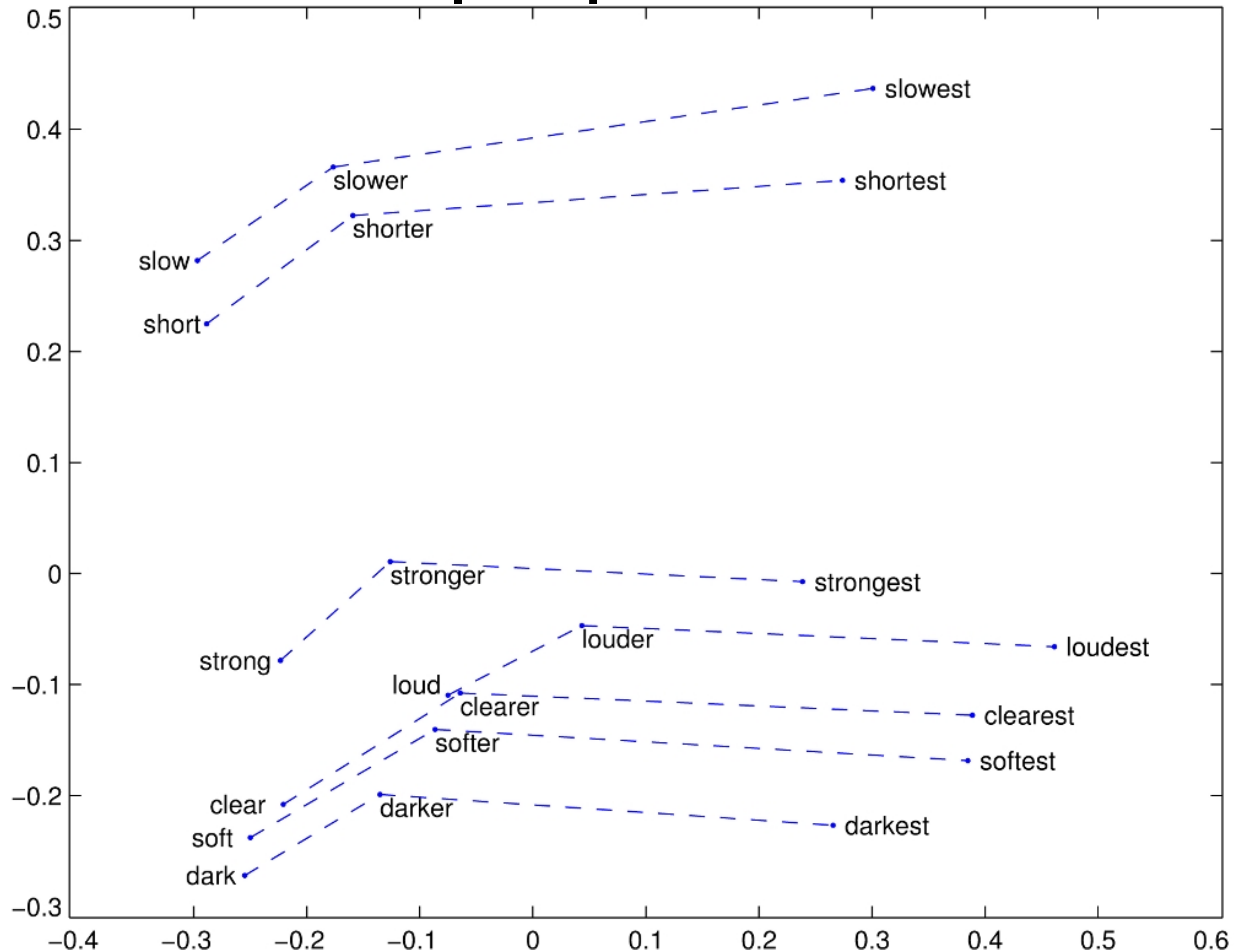
# Linear semantic properties?

<https://nlp.stanford.edu/projects/glove/>



# Linear semantic properties?

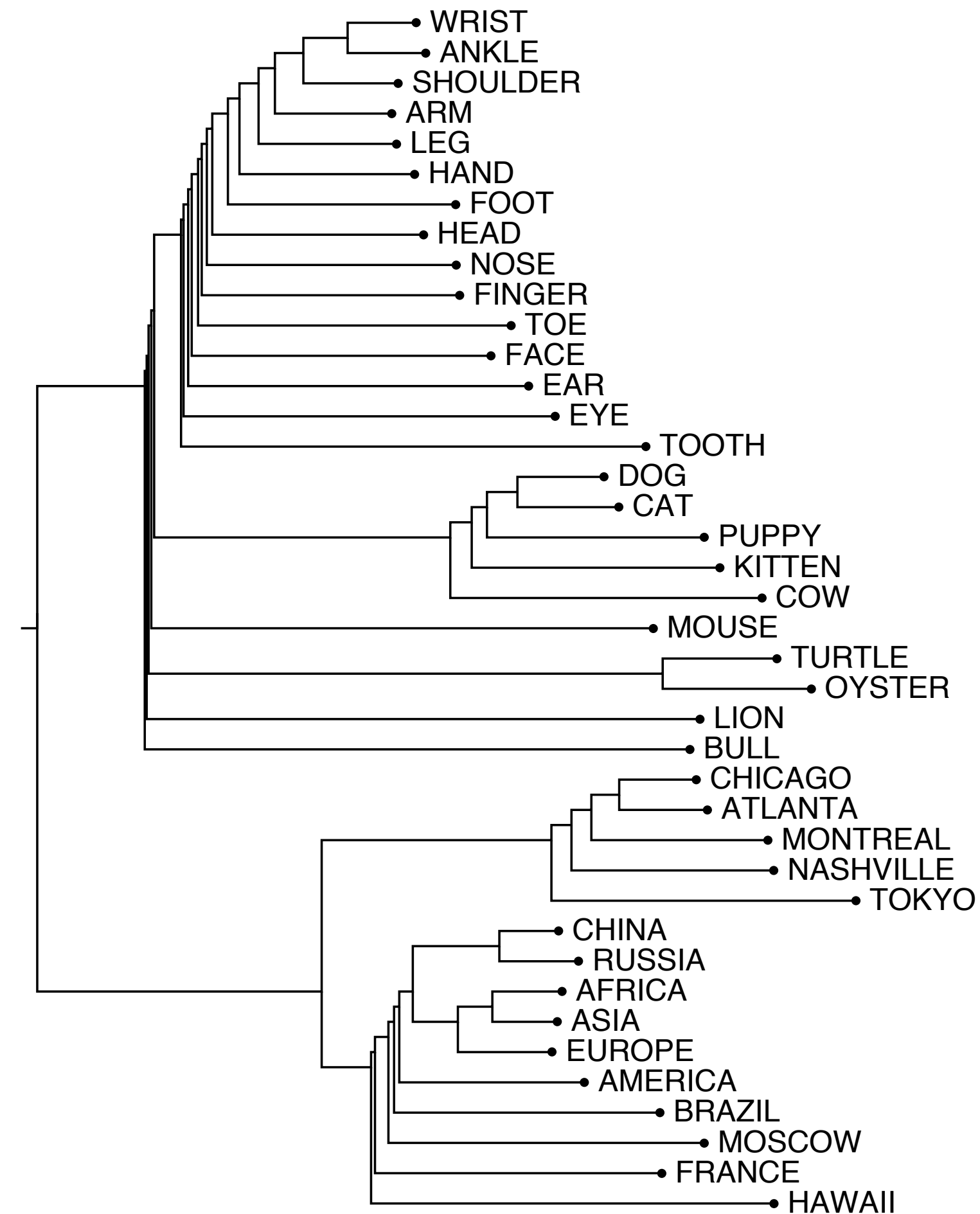
<https://nlp.stanford.edu/projects/glove/>



# Linear semantic properties?

- Linear? Sometimes!
  - Though cool early examples on analogical structure didn't always totally pan out
- There definitely is *some* semantic structure in distributional word embeddings, in any case

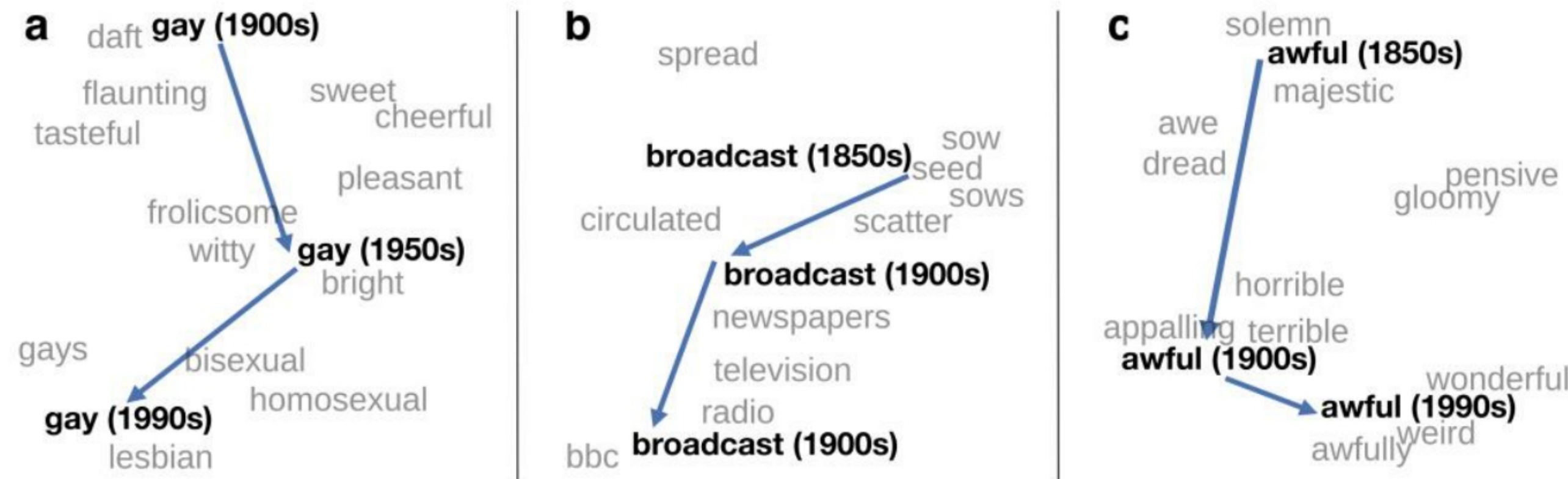
# embeddings may have larger-scale semantic structure?



# embeddings may have larger-scale semantic structure?

- Hierarchical distributional word clusters, trained from tweets:  
[http://www.cs.cmu.edu/~ark/TweetNLP/  
cluster\\_viewer.html](http://www.cs.cmu.edu/~ark/TweetNLP/cluster_viewer.html)
- What distinctions is it learning?

# Training corpus is crucial



Word use over time  
[Hamilton et al. 2016]

- Train on general corpus or domain-specific?
  - Historical periods
  - Particular subreddits
  - etc. etc.
- The content of the pretraining corpus is very important!!  
Datasets should be documented.

# ok so what can we do with them?

- *Transfer learning* from large, unsup. corpus
- Document embeddings
  - 1. Supervised learning: Bag-of-Embeddings logreg
    - labeled train docs->labeled new docs
  - 2. Unsupervised learning / exploratory analysis
    - docs->[analysis]
- Wordlist-based inferences
  - 3. Semi-automatic dictionary expansion
    - (words->words)
  - 4. DDR: Distrib. Dict. Representations
    - (words->docs)

# (1) Sup. learning with document embedding

- Instead of bag-of-words, can we derive a latent embedding of a document/sentence?
  - "Bag of embeddings" or "averaged word embeddings" representation
  - You can use it just like a BOW logistic regression - it's just a different type of feature vector
  - Pros/cons?
- Especially for shorter texts, BoE LR typically outperforms BOW LR.

## (2) Unsup. learning with document embedding

- Example: tweets about mass shootings ([Demszky et al. 2019](#))
  1. Average word embeddings => tweet embeddings
  2. Cluster tweets (k-means)
  3. Interpret clusters' words (closest to centroid)

| Topic                               | 10 Nearest Stems                                                                            |
|-------------------------------------|---------------------------------------------------------------------------------------------|
| news<br>(19%)                       | break, custodi, #breakingnew, #updat, confirm, fatal, multipl, updat, unconfirm, sever      |
| investigation<br>(9%)               | suspect, arrest, alleg, apprehend, custodi, charg, accus, prosecutor, #break, ap            |
| shooter's identity & ideology (11%) | extremist, radic, racist, ideolog, label, rhetor, wing, blm, islamist, christian            |
| victims & location<br>(4%)          | bar, thousand, california, calif, among, los, southern, veteran, angel, via                 |
| laws & policy<br>(14%)              | sensibl, regul, requir, access, abid, #gunreformnow, legisl, argument, allow, #guncontolnow |
| solidarity<br>(13%)                 | affect, senseless, ach, heart, heartbroken, sadden, faculti, pray, #prayer, deepest         |
| remembrance<br>(6%)                 | honor, memori, tuesday, candlelight, flown, vigil, gather, observ, honour, capitol          |
| other<br>(23%)                      | dude, yeah, eat, huh, gonna, ain, shit, ass, damn, guess                                    |

Table 1: Our eight topics (with their average proportions across events) and nearest-neighbor stem embeddings to the cluster centroids. Topic names were manually assigned based on inspecting the tweets.

# Application: keyword expansion

- I have a few keywords for my task. Are there any I missed?
- QUERY: keyword => nearest neighbors
- Automated or semi-automatedly acquire new terms from embedding neighbors
  - Much better than dreaming up terms yourself!
- Other non-embedding lexical resources can do this too (e.g. WordNet), but word embeddings typically cover a *lot* of diverse vocabulary, with implicit knowledge from whatever the corpus is

- stopped here 9/30

# Application: doc sim to words

- Given a word list to represent a concept, can we score a document for how much it expresses that concept?
- Count based approach?

# Application: doc sim to words

- Given a word list to represent a concept, can we score a document for how much it expresses that concept?
- DDR is a very simple embedding approach:
  - Average the word lists embeddings to create a concept vector
  - Average a doc's words to create a document vector
  - Apply cosine similarity!
- Supplying a set of keywords is *low-supervision*, or low-expertise, approach compared to labeling docs
  - Though you don't get a nice logreg probability (until you label some...)

# LIWC "posemo" list

Input:

accept, accepta\*, accepted, accepting, accepts, active\*, admir\*, ador\*, advantag\*, adventur\*, affection\*, agree, agreeab\*, agreed, agreeing, agreement\*, agrees, alright\*, amaz\*, amor\*, amus\*, aok, appreciat\*, assur\*, attachment\*, attract\*, award\*, awesome, beaut\*, beloved, benefic\*, benefit, benefits, benefitt\*, benevolen\*, benign\*, best, better, bless\*, bold\*, bonus\*, brave\*, bright\*, brillian\*, calm\*, care, cared, carefree, careful\*, cares, caring, casual, casually, certain\*, challeng\*, champ\*, charit\*, charm\*, cheer\*, cherish\*, chuckl\*, clever\*, comed\*, comfort\*, commitment\*, compassion\*, compliment\*, confidence, confident, confidently, considerate, contented\*, contentment, convinc\*, cool, courag\*, create\*, creati\*, credit\*, cute\*, cutie\*, daring, darlin\*, dear\*, definite, definitely, delectabl\*, delicate\*, delicious\*, deligh\*, determina\*, determined, devot\*, digni\*, divin\*, dynam\*, eager\*, ease\*, easie\*, easily, easiness, easing, easy\*, ecsta\*, efficien\*, elegan\*, encourag\*, energ\*, engag\*, enjoy\*, entertain\*, enthus\*, excel\*, excit\*, fab, fabulous\*, faith\*, fantastic\*, favor\*, favour\*, fearless\*, festiv\*, fiesta\*, fine, flatter\*, flawless\*, flexib\*, flirt\*, fond, fondly, fondness, forgave, forgiv\*, free, freeb\*, freed\*, freeing, freely, freeness, freer, frees\*, friend\*, fun, funn\*, genero\*, gentle, gentler, gentlest, gently, giggl\*, giver\*, giving, glad, gladly, glamor\*, glamour\*, glori\*, glory, good, goodness, gorgeous\*, grace, graced, graceful\*, graces, graci\*, grand, grande\*, gratef\*, grati\*, great, grin, grinn\*, grins, ha, haha\*, handsom\*, happi\*, happy, harmless\*, harmon\*, heartfelt, heartwarm\*, heaven\*, heh\*, helper\*, helpful\*, helping, helps, hero\*, hilarious, hoho\*, honest\*, honor\*, honour\*, hope, hoped, hopeful, hopefully, hopefulness, hopes, hoping, hug, hugg\*, hugs, humor\*, humour\*, hurra\*, ideal\*, importan\*, impress\*, improve\*, improving, incentive\*, innocen\*, inspir\*, intell\*, interest\*, invigor\*, joke\*, joking, joll\*, joy\*, keen\*, kidding, kind, kindly, kindn\*, kiss\*, laidback, laugh\*, libert\*, like, likeab\*, liked, likes, liking, livel\*, lmao, lol, love, loved, lovely, lover\*, loves, loving\*, loyal\*, luck, lucked, lucki\*, lucks, lucky, madly, magnific\*, merit\*, merr\*, neat\*, nice\*, nurtur\*, ok, okay, okays, oks, openminded\*, openness, opport\*, optimal\*, optimi\*, original, outgoing, painl\*, palatabl\*, paradise, partie\*, party\*, passion\*, peace\*, perfect\*, play, played, playful\*, playing, plays, pleasant\*, please\*, pleasing, pleasur\*, popular\*, positiv\*, prais\*, precious\*, prettie\*, pretty, pride, privileg\*, prize\*, profit\*, promis\*, proud\*, radian\*, readiness, ready, reassur\*, relax\*, relief, reliev\*, resolv\*, respect, revigor\*, reward\*, rich\*, rofl, romanc\*, romantic\*, safe\*, satisf\*, save, scrumptious\*, secur\*, sentimental\*, share, shared, shares, sharing, silli\*, silly, sincer\*, smart\*, smil\*, sociab\*, soulmate\*, special, splend\*, strength\*, strong\*, succeed\*, success\*, sunnier, sunniest, sunny, sunshin\*, super, superior\*, support, supported, supporter\*, supporting, supportive\*, supports, suprem\*, sure\*, surpris\*, sweet, sweetheart\*, sweetie\*, sweetly, sweetness\*, sweets, talent\*, tehe, tender\*, terrific\*, thank, thanked, thankf\*, thanks, thoughtful\*, thrill\*, toleran\*, tranquil\*, treasur\*, treat, triumph\*, true, trueness, truer, truest, truly, trust\*, truth\*, useful\*, valuabl\*, value, valued, values, valuing, vigor\*, vigour\*, virtue\*, virtuo\*, vital\*, warm\*, wealth\*, welcom\*, well, win, winn\*, wins, wisdom, wise\*, won, wonderf\*, worship\*, worthwhile, wow\*, yay, yays

NNs of its  
semantic centroid:



Fig. 4 Nearest neighbors of the LIWC positive emotions dictionary



