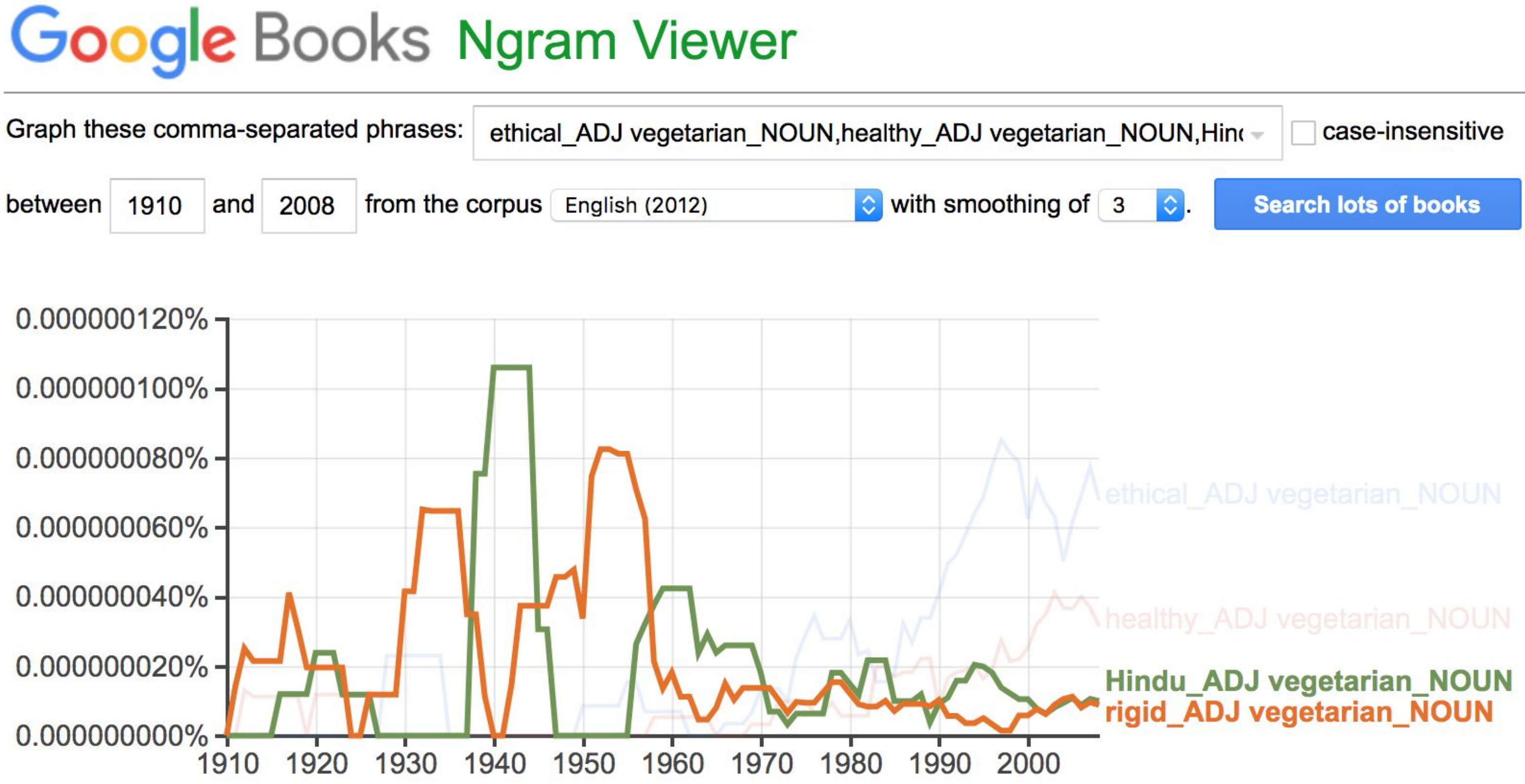
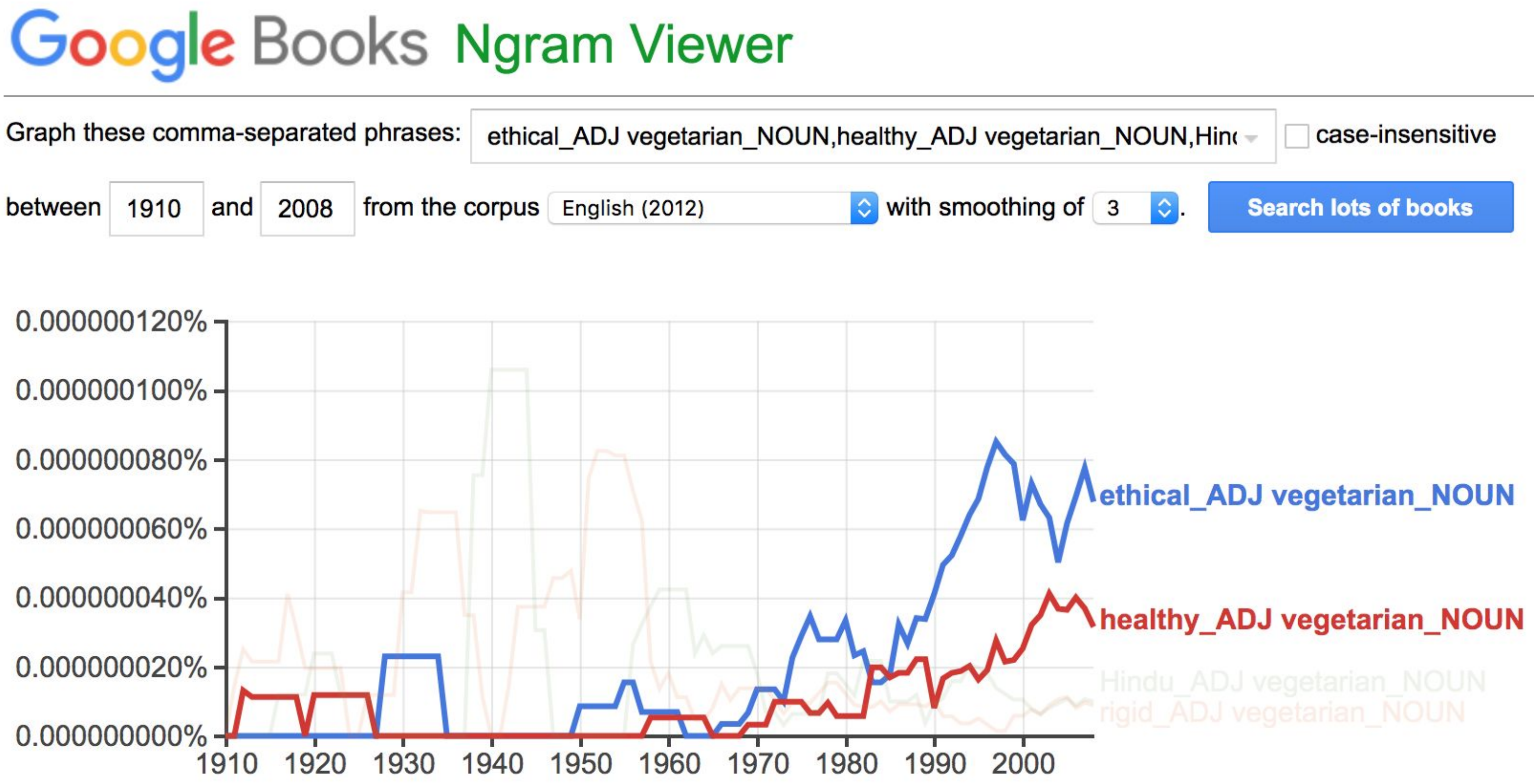
Chaining Algorithms and Adjective Extension

Karan Grewal
Language, Cognition, and Computation Group
April 24, 2020

Changes in adjective use



Changes in adjective use



Changes in adjective use

are changes in adjective use random, or is there a computational underpinning?

small-scale studies suggest there is regularity:

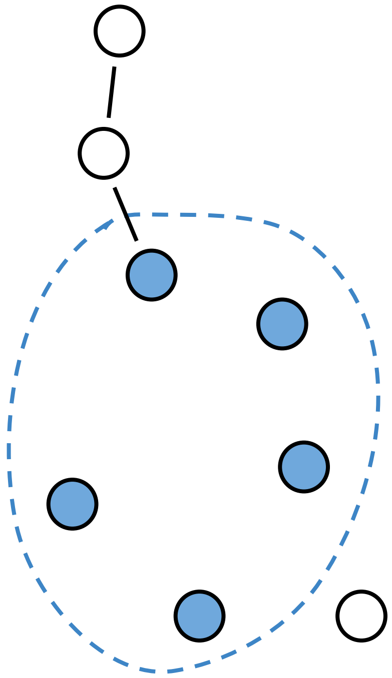
color to sound	*clear* blue → *clear* voice, *faint* shade → *faint* scream
touch to taste	*sharp* edge → *sharp* taste

Synaesthetic adjectives (Williams, 1976)

Semantic chaining

Lakoff (and others) proposed that semantic categories grow through chaining → attracting new stimuli based on semantic similarity

chaining can take various forms: e.g. nearest neighbor growth

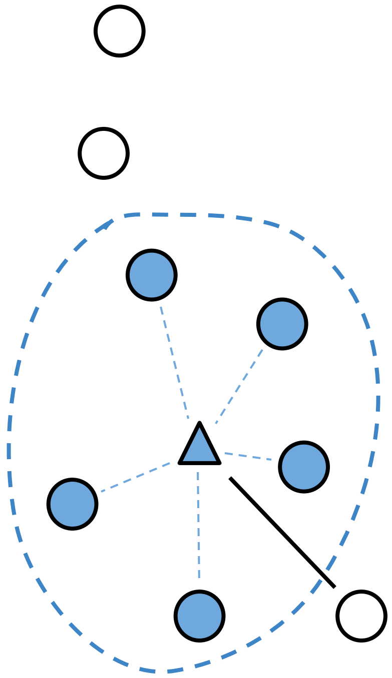


Semantic chaining (Lakoff, 1987; Malt et al., 1999)

Semantic chaining

Lakoff (and others) proposed that semantic categories grow through chaining → attracting new stimuli based on semantic similarity

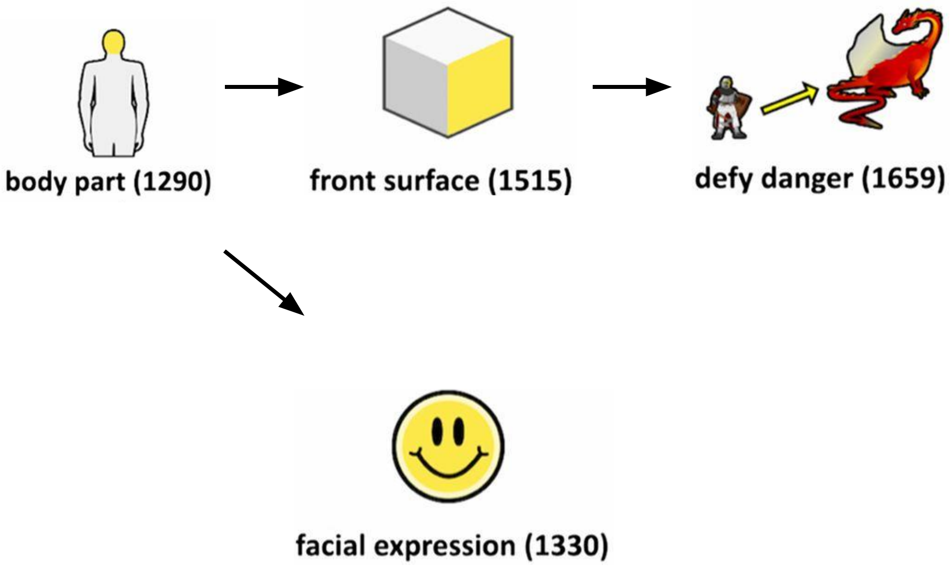
chaining can take various forms: e.g. nearest neighbor growth, <u>centroid growth</u>



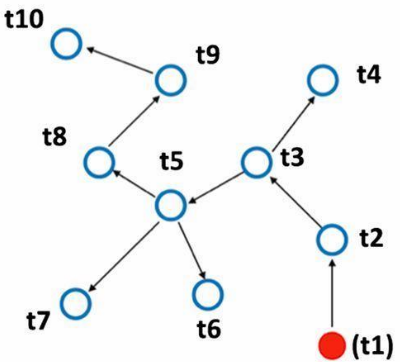
Semantic chaining (Lakoff, 1987; Malt et al., 1999)

Computational studies of semantic chaining

"face"



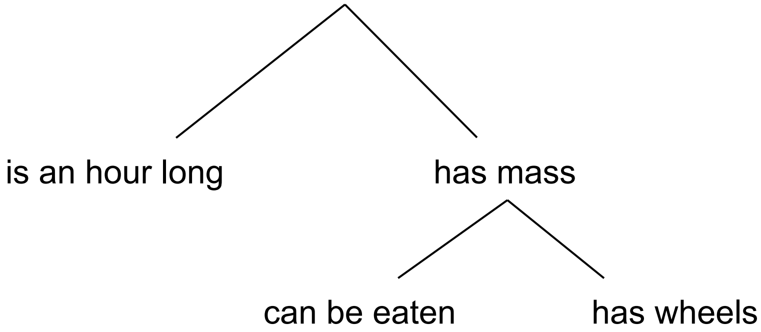
a minimal spanning tree algorithm best reconstructs the extension of word senses



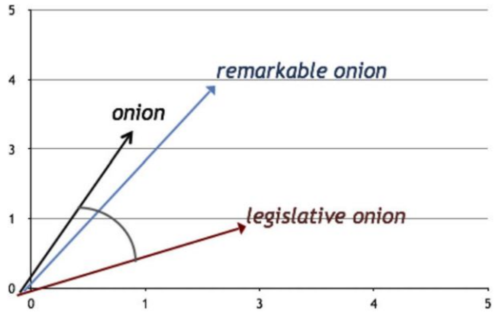
Chaining & emergence of word senses (Ramiro et al., 2018)

Computational work on acceptable adjective-noun pairs

(1) ontological constraints determine plausibility



(2) vector space models can capture plausibility



these approaches are "static" with respect to time!

Ontological constraints (Schmidt et al., 2006),
Sensible compositions via vector space models (Vecchi et al., 2016)

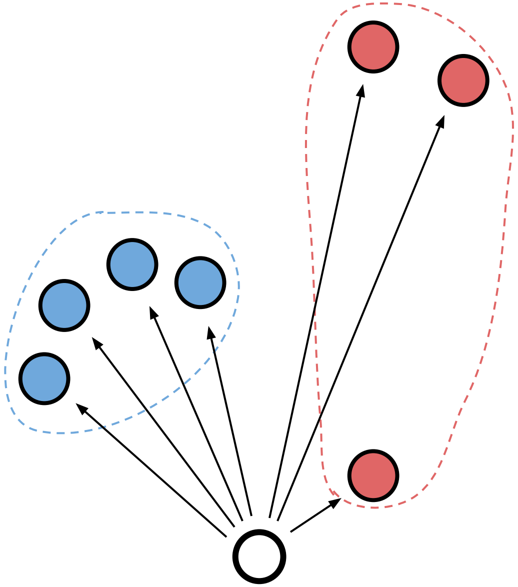
Chaining and adjective extension

given noun $n*$ and information up to time t about its (a) semantics, and (b) past adjective pairings, want to predict which adjectives it will pair with at time $t + \Delta$

approach:

- treat adjectives as categories, and place nouns into 1 or more categories
- define a likelihood function $p(n* \mid a)$ based on chaining that models extension

Exemplar model



likelihood:

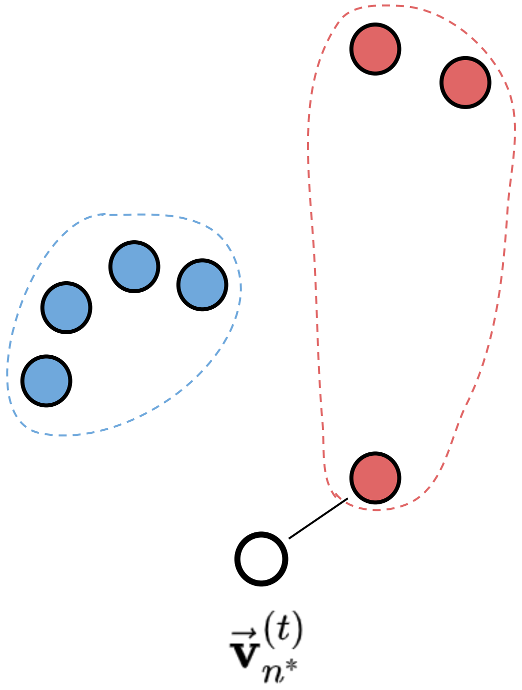
$$p(n^*|a) \propto \frac{1}{\left|\{n\}_a^{(t)}\right|} \sum_{n \in \{n\}_a^{(t)}} \mathrm{sim}\left(\vec{\mathbf{v}}_{n^*}^{(t)}, \vec{\mathbf{v}}_n^{(t)}\right)$$

where the similarity between nouns $n*$ and n is

$$\mathrm{sim}\left(\vec{\mathbf{v}}_{n^*}^{(t)}, \vec{\mathbf{v}}_n^{(t)}\right) = e^{-d\left(\vec{\mathbf{v}}_{n^*}^{(t)}, \vec{\mathbf{v}}_n^{(t)}\right)^2}$$

Exemplar theory (Nosofsky, 1986)

k-nearest neighbors (k-NN) model



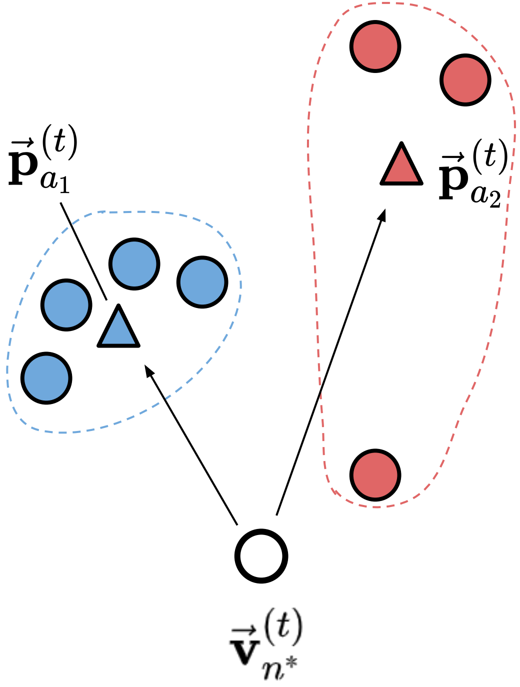
find n*'s k closest nouns, and count their "votes"

$$p(n^*|a) \propto \frac{1}{\left|\{n\}_a^{(t)}\right|} \sum_{j=1}^{k} \mathbb{1}\left[n_j \in \{n\}_a^{(t)}\right]$$

where j indexes the k closest nouns to n*

Nearest neighbor chaining (Sloman et al., 2001; Xu et al., 2016)

Prototype model



prototype for adjective a:

$$\vec{\mathbf{p}}_a^{(t)} = \mathbb{E}\left[n \in \{n\}_a^{(t)}\right]$$

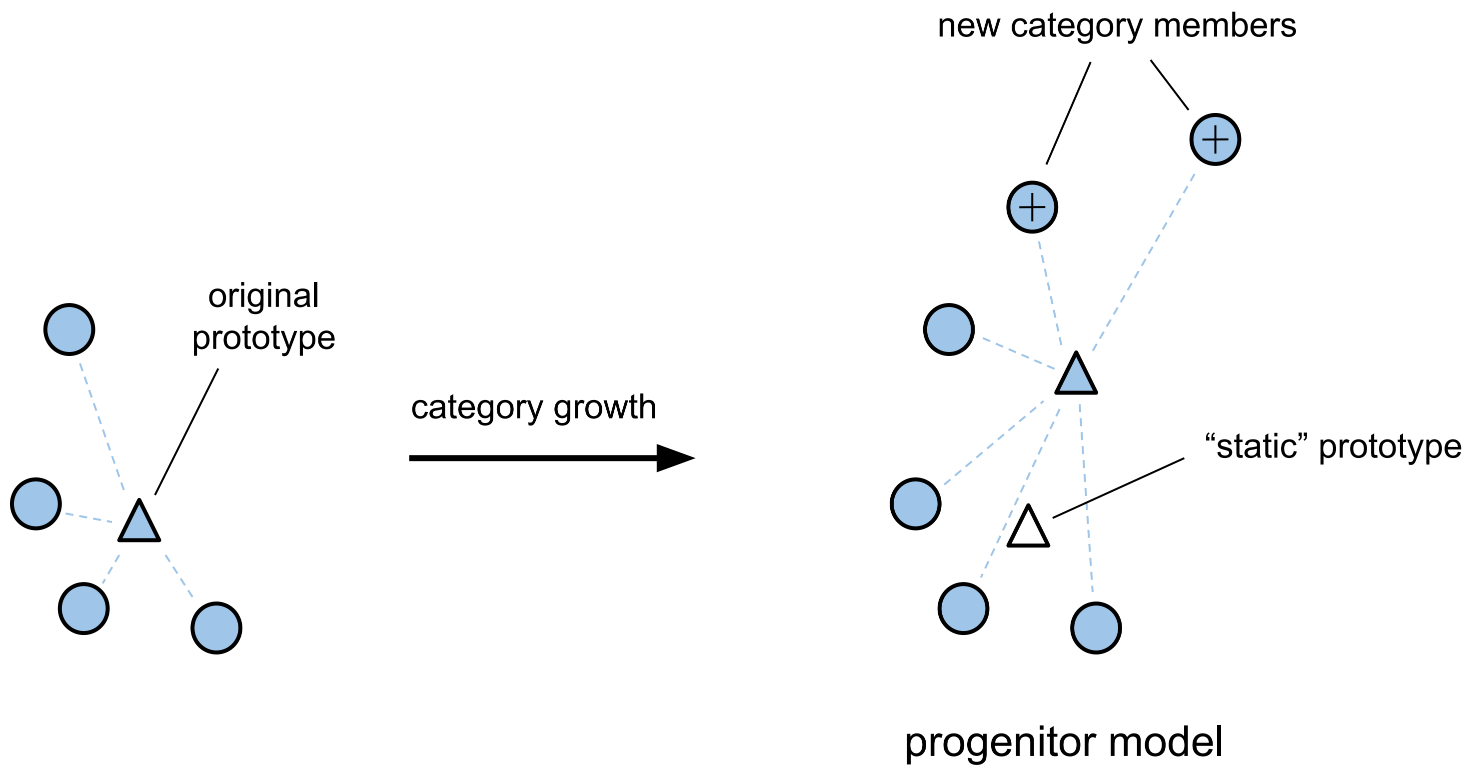
$$\approx \frac{1}{\left|\{n\}_a^{(t)}\right|} \sum_{n \in \{n\}_a^{(t)}} \vec{\mathbf{v}}_n^{(t)}$$

likelihood:

$$p(n^* | a) \propto \mathrm{sim}(\vec{\mathbf{v}}_{n^*}^{(t)}, \vec{\mathbf{p}}_a^{(t)})$$

Prototype theory (Rosch, 1975)

Prototype model



Prototype theory (Rosch, 1975)

Probabilistic framework

$$p(a|n^*)^{(t+\Delta)} \propto p(n^*|a)^{(t)} \underbrace{p(a)^{(t)}}$$

$$\propto \left| \{n\}_a^{(t)} \right|$$

size-based prior
(according to corpus statistics)

Probabilistic framework

add kernel parameter into similarity function: $$\mathrm{sim}(\cdot,\cdot) = \exp\left(-\frac{d(\cdot,\cdot)^2}{h}\right)$$

exemplar:

$$p(n^*|a) \propto \sum_{n\in\{n\}_a^{(t)}} \mathrm{sim}\left(\vec{\mathbf{v}}_{n^*}^{(t)}, \vec{\mathbf{v}}_n^{(t)}\right)$$
$$\propto \sum_{n\in\{n\}_a^{(t)}} \exp\left(-\frac{d(\vec{\mathbf{v}}_{n^*}^{(t)}, \vec{\mathbf{v}}_n^{(t)})}{h}\right)$$

prototype:

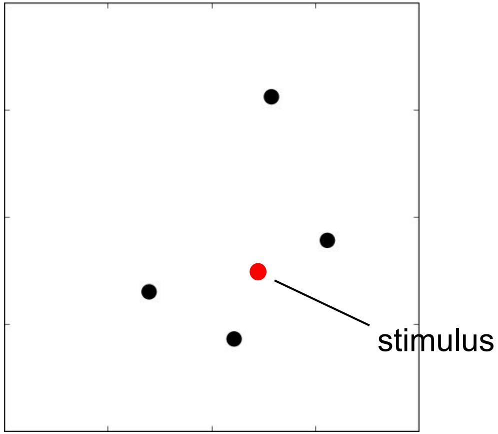
$$p(n^*|a) \propto \mathrm{sim}\left(\vec{\mathbf{v}}_{n^*}^{(t)}, \vec{\mathbf{p}}_a^{(t)}\right)$$
$$\propto \exp\left(-\frac{d(\vec{\mathbf{v}}_{n^*}^{(t)}, \vec{\mathbf{p}}_a^{(t)})^2}{h}\right)$$

$$p(n^*|a) \propto \sum_{n\in\{n\}_a^{(t)}} \exp\left(-\frac{d(\vec{\mathbf{v}}_{n^*}^{(t)},\vec{\mathbf{v}}_n^{(t)})}{h}\right)$$

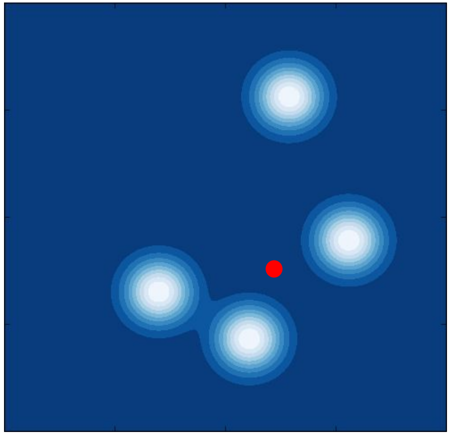
Probabilistic framework

exemplar likelihood with kernel parameter is equivalent to mixture-of-Gaussians

h controls how many neighbors the exemplar model “pays attention” to

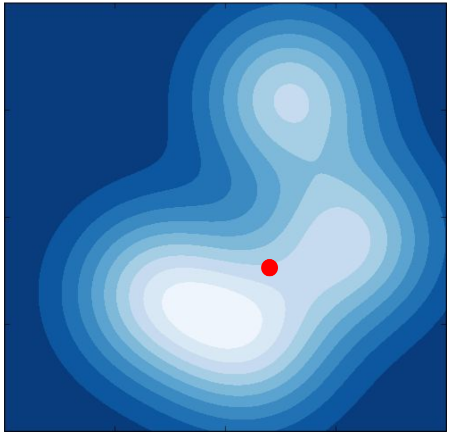


exemplars



small h

peaked density function



big h

wider density function

Data and materials

extracted all adjective-noun co-occurrences (according to POS tags) and their timestamps from the Google books English-All corpus; 200 years worth of data (1800-2000)

"ethical_ADJ vegetarian_NOUN" (t = 1969, count = 4)
"ethical_ADJ vegetarian_NOUN" (t = 1970, count = 11)
⋮

grouped all co-occurrences into bins of size Δ = 10 years

In total, extracted co-occurrences for 67k nouns and 14k adjectives

Data and materials

how to deal with semantics only up to time t? → diachronic word embeddings

Diachronic word embeddings (Hamilton et al., 2016)

Adjectives

tested our models on 3 sets of adjectives, two of which are (1) frequent adjectives, and (2) random adjectives

constructed (1) and (2) by clustering adjective embeddings (Word2Vec), then sampling from each cluster via frequency-based sampling and uniform sampling resp.

cultures		personal traits		touch perceptions	
Frequent	Random	Frequent	Random	Frequent	Random
Asian	*Hungarian*	*polite*	*chatty*	*warm*	*chilly*
Christian	*Thai*	*intelligent*	*unorthodox*	*dense*	*watery*
American	*Cornish*	*passionate*	*amiable*	*dry*	*fertile*
European	*Catalan*	*energetic*	*communicative*	*tropical*	*drizzling*

Adjectives

… and (3) synaesthetic adjectives, a small group of adjectives that transfer to new domains and have been studied extensively

touch to color	*dull*, *light*, *warm*
color to sound	*bright*, *brilliant*, *faint*, *light*, *vivid*

Synaesthetic adjectives (Williams, 1976)

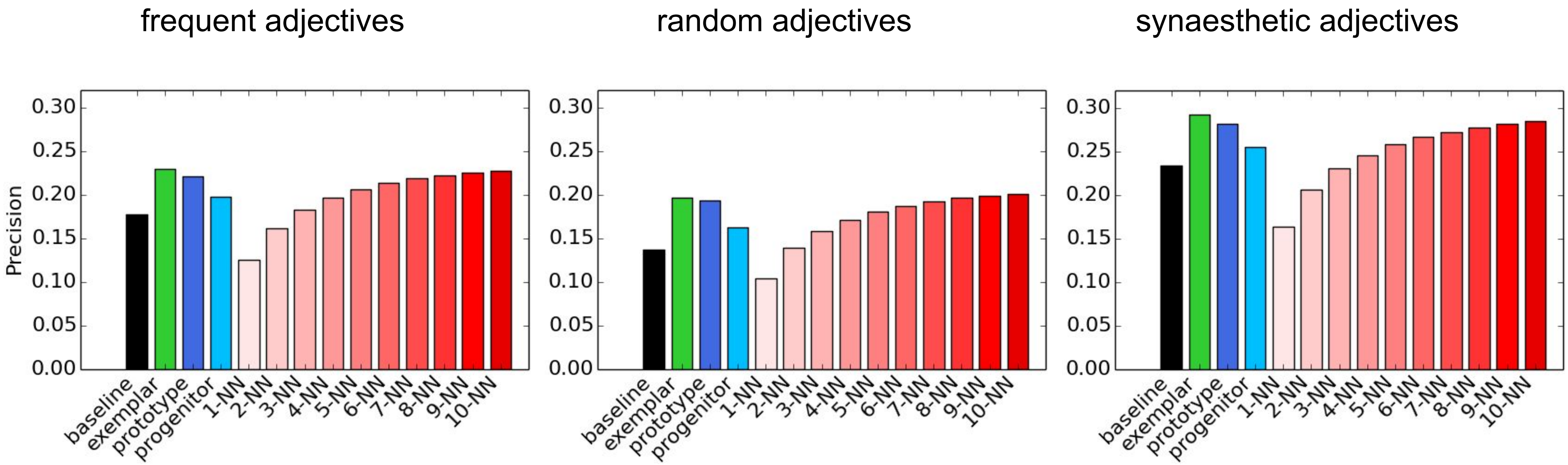
Model evaluation

If noun n^* co-occurred with m new adjectives at time $t + \Delta$, find the top m adjectives with the highest posterior probability that previously did not pair with n^* as the retrieved positives

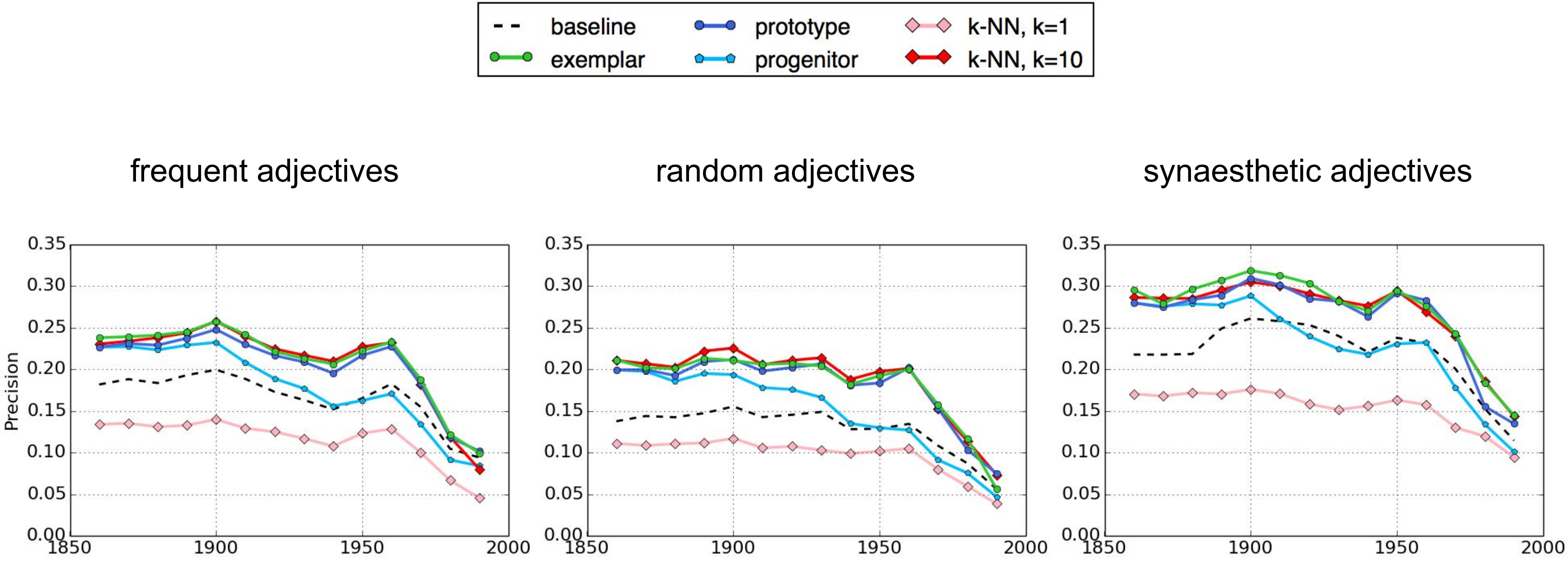
$$\text{precision} = \frac{\sum \text{total correct adjective predictions at time } t + \Delta}{\sum \text{total predictions at time } t + \Delta}$$

also optimize precision to learn the kernel parameter for the exemplar and prototype models

Results (aggregate)



Results (per decade)



Results (predictions)

alcohol, 1920s

new adjectives	*female, analogous, red, bitter, marked, illegal*
baseline prediction	*perfect, extraordinary, moral, physical, western, christian (0/6)*
exemplar prediction	***red**, moral, artificial, dense, perfect, **marked** (2/6)*
prototype prediction	*artificial, perfect, **marked**, **red**, physical, moral (2/6)*
10-NN prediction	***red**, moral, dense, perfect, **analogous**, artificial (2/6)*

Vietnam, 1960s

new adjectives	*western, tropical, eastern, colonial, particular, more, top, poor, American*
baseline prediction	*same, **more**, great, **particular**, **American**, different, natural, human, English (3/9)*
exemplar prediction	***western**, **eastern**, **more**, **particular**, great, **colonial**, inner, same, **poor** (6/9)*
prototype prediction	*great, same, **western**, **more**, **American**, **eastern**, **particular**, European, French (5/9)*
10-NN prediction	***western**, **eastern**, **more**, **tropical**, **colonial**, great, better, inner, **particular** (6/9)*

Results (predictions)

cigarette, 1880s

new adjectives	*better, modern, several, excessive, American, social*
baseline prediction	*original, particular, English, natural, perfect,* ***modern*** *(1/6)*
exemplar prediction	*black, red, English, poor, original, particular (0/6)*
prototype prediction	*red, black, dry, warm, cold, English (0/6)*
10-NN prediction	*original, warm, particular, red, English, dry (0/6)*

cigarette, 1920s

new adjectives	*different, odd, worn, scattered, illegal, wrong*
baseline prediction	*natural,* ***different****, sufficient, extraordinary, moral, mental (1/6)*
exemplar prediction	***different****, natural, warm, sufficient, solid, inner (1/6)*
prototype prediction	*warm,* ***different****, top, natural, solid, circular (1/6)*
10-NN prediction	*natural, top, warm,* ***different****, sufficient, conventional (1/6)*

Conclusions

our results suggest that semantic chaining largely explains adjective extension as our predictive models consistently perform better than the size-based prior across 3 adjectives sets

future work: does information in the visual world also explain the evolution of adjective-noun pairs?