# CROSS-CUTTING MODELS OF DISTRIBUTIONAL LEXICAL SEMANTICS

Joseph S. Reisinger
The University of Texas at Austin

Doctoral Dissertation Proposal

June 17th, 2010

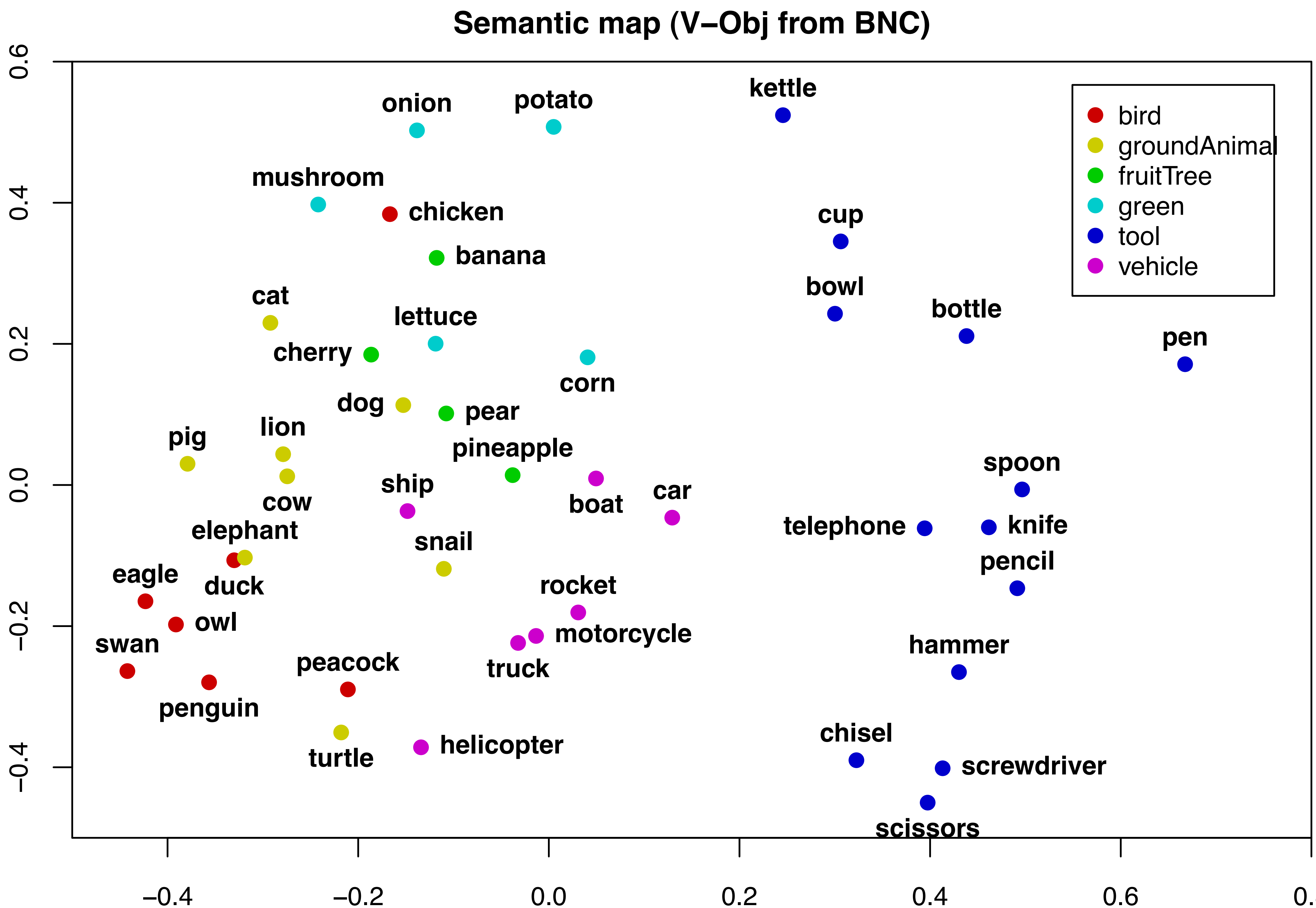
# Lexical semantics

- Can we infer / represent the meaning of words?
- Knowledge-based approaches (WordNet, FrameNet, etc)
  - rich structure, e.g. directed-acyclic synset graph
  - hand-built, limited to a few languages
- Distributional approaches ("*you shall know a word by the company it keeps*" Firth 1957)
  - more scalable, depends only on raw corpora
  - less rich categorical structure

- S: (n) **lynx** (a text browser)
- S: (n) **lynx**, catamount (short-tailed wildcats with usually tufted ears; valued for their fur)
  - *direct hyponym* / ***full hyponym***
    - S: (n) common lynx, Lynx lynx (of northern Eurasia)
    - S: (n) Canada lynx, Lynx canadensis (of northern North America)
    - S: (n) bobcat, bay lynx, Lynx rufus (small lynx of North America)
    - S: (n) spotted lynx, Lynx pardina (of southern Europe)
    - S: (n) caracal, desert lynx, Lynx caracal (of deserts of northern Africa and southern Asia)
  - *member holonym*
    - S: (n) genus Lynx (lynxes)
  - *direct hypernym* / *inherited hypernym* / *sister term*
    - S: (n) wildcat (any small or medium-sized cat resembling the domestic cat and living in the wild)
      - S: (n) cat, true cat (feline mammal usually having thick soft fur and no ability to roar: domestic cats; wil
        - S: (n) feline, felid (any of various lithe-bodied roundheaded fissiped mammals, many with retractile
          - S: (n) carnivore (a terrestrial or aquatic flesh-eating mammal) *"terrestrial carnivores have fou limb"*
            - S: (n) placental, placental mammal, eutherian, eutherian mammal (mammals having a pla monotremes and marsupials)
              - S: (n) mammal, mammalian (any warm-blooded vertebrate having the skin more or born alive except for the small subclass of monotremes and nourished with milk)
                - S: (n) vertebrate, craniate (animals having a bony or cartilaginous skeleton wit a large brain enclosed in a skull or cranium)
                  - S: (n) chordate (any animal of the phylum Chordata having a notochord c
                    - S: (n) animal, animate being, beast, brute, creature, fauna (a living or movement)
                      - S: (n) organism, being (a living thing that has (or can develop) 

# Lexical semantics

- Can we infer / represent the meaning of words?
- Knowledge-based approaches (WordNet, FrameNet, etc)
  - rich structure, e.g. directed-acyclic synset graph
  - hand-built, limited to a few languages
- Distributional approaches (“*you shall know a word by the company it keeps*” Firth 1957)
  - more scalable, depends only on raw corpora
  - less rich categorical structure


Semantic map (V−Obj from BNC)
bird
groundAnimal
fruitTree
green
tool
vehicle
onion
potato
kettle
mushroom
chicken
banana
cup
bowl
bottle
pen
cat
lettuce
cherry
corn
dog
pear
pig
lion
pineapple
spoon
ship
boat
car
cow
telephone
knife
elephant
snail
pencil
eagle
duck
rocket
owl
motorcycle
swan
hammer
peacock
truck
penguin
chisel
helicopter
screwdriver
turtle
scissors
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8

(Evert 2010)

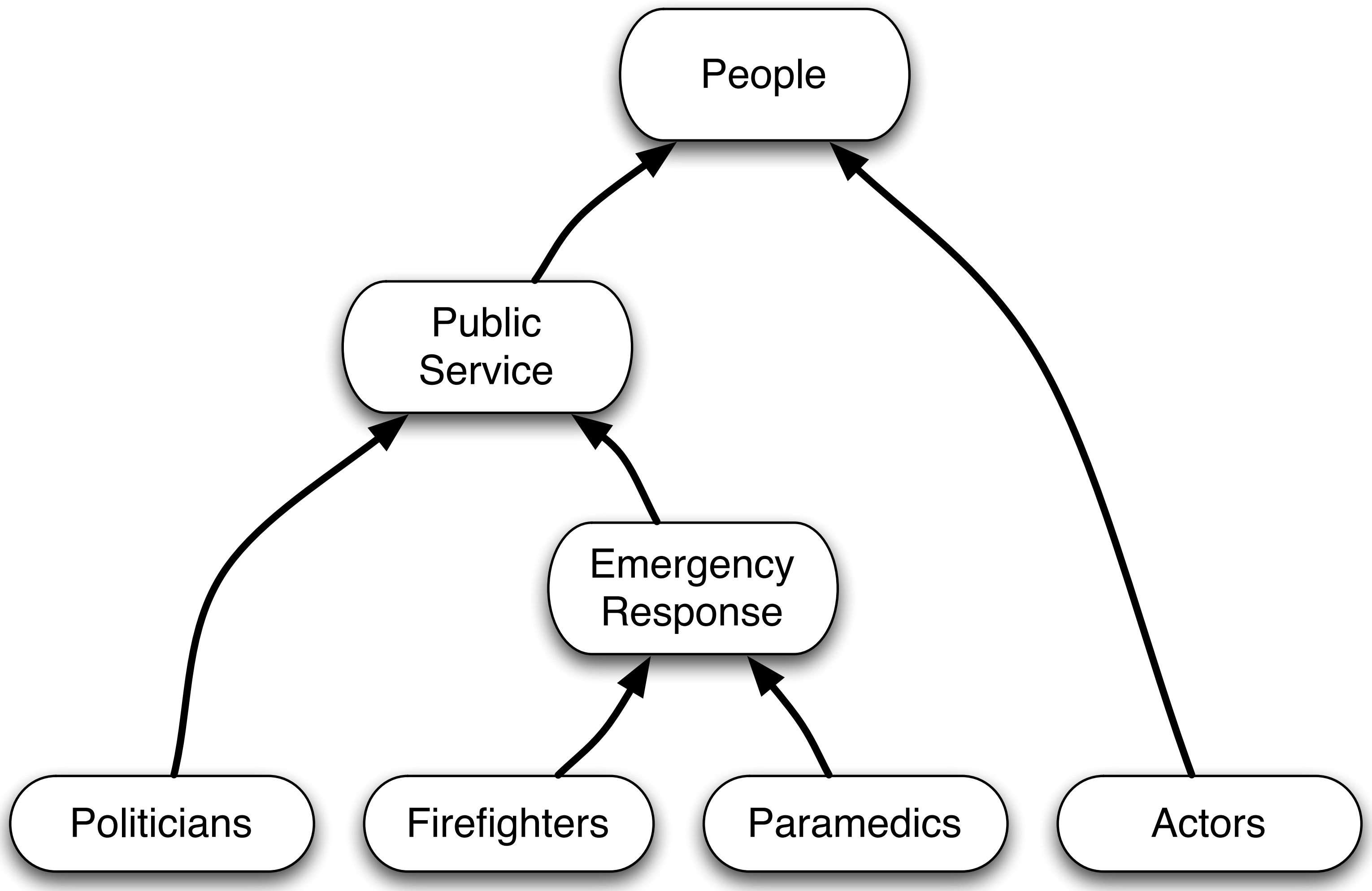
# Lexical semantics

- Can we infer / represent the meaning of words?
- Knowledge-based approaches (WordNet, FrameNet, etc)
  - rich structure, e.g. directed-acyclic synset graph
  - hand-built, limited to a few languages
- Distributional approaches ("*you shall know a word by the company it keeps*" Firth 1957)
  - more scalable, depends only on raw corpora
  - less rich categorical structure

# A hypothesis

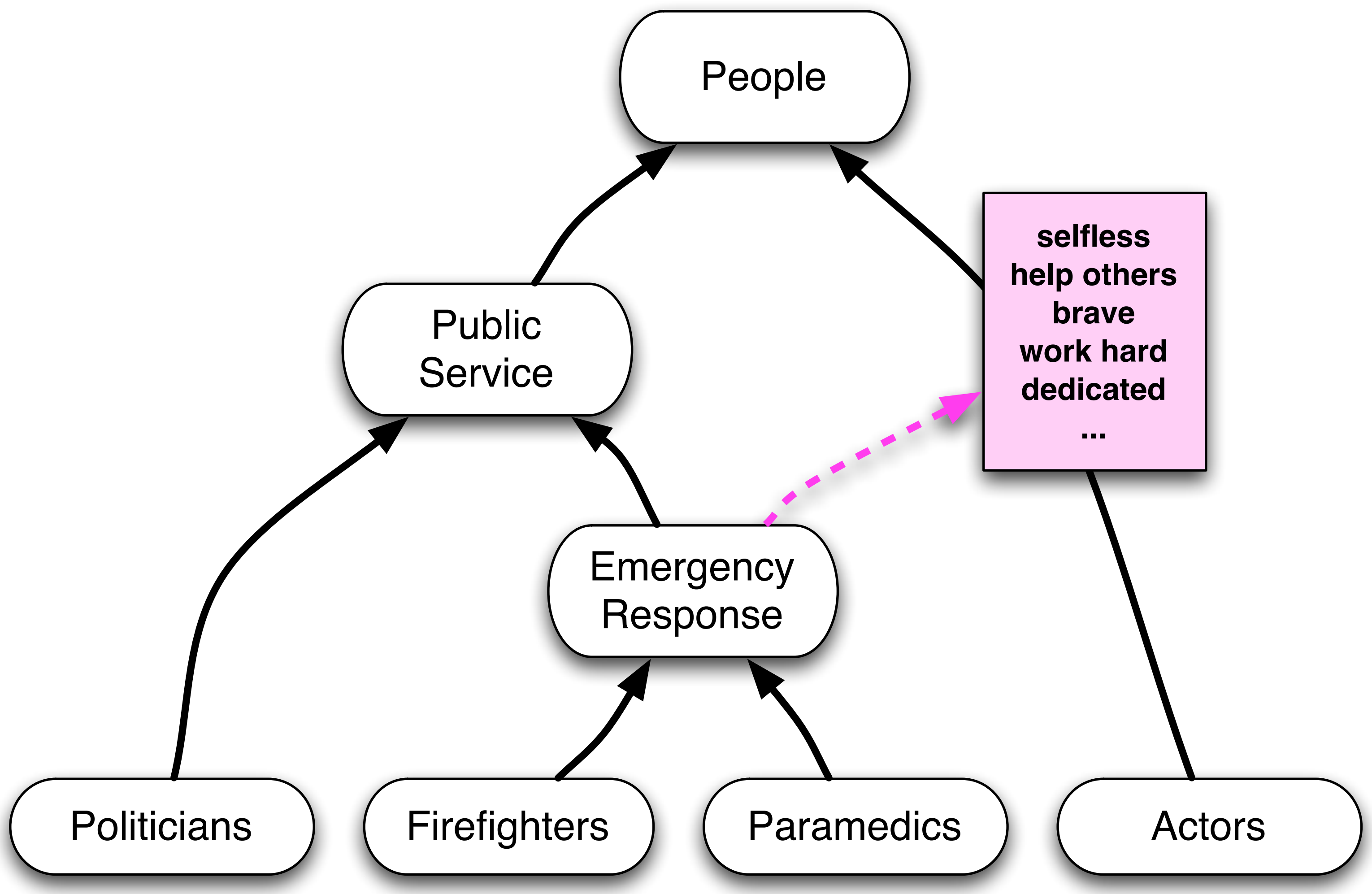
- Hand-built knowledge bases like Wikipedia and WN are far richer structurally than distributional models
- Lexical organization is driven by human conceptual organization

# A hypothesis



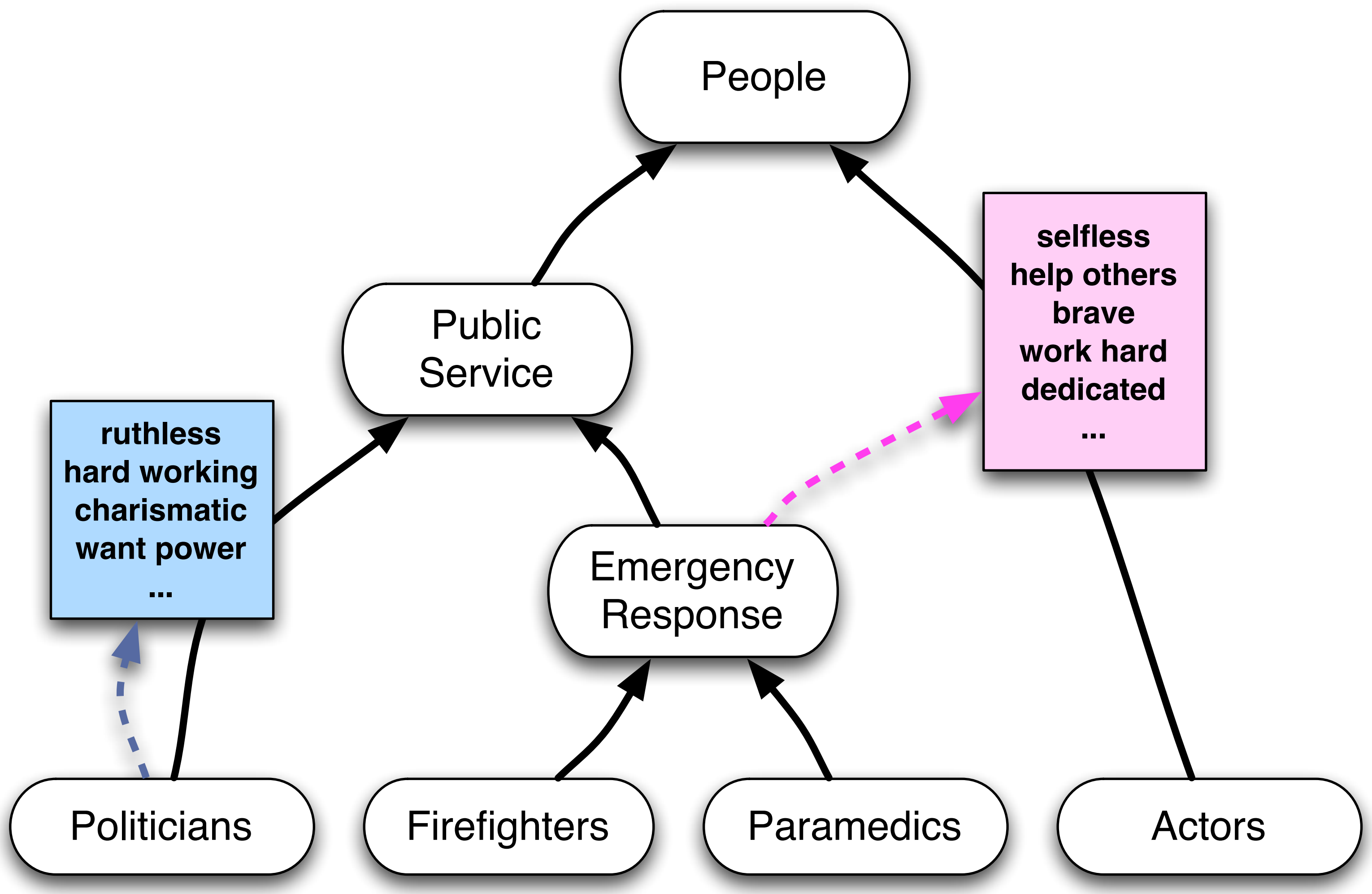

a reasonable hierarchical clustering of people

# A hypothesis



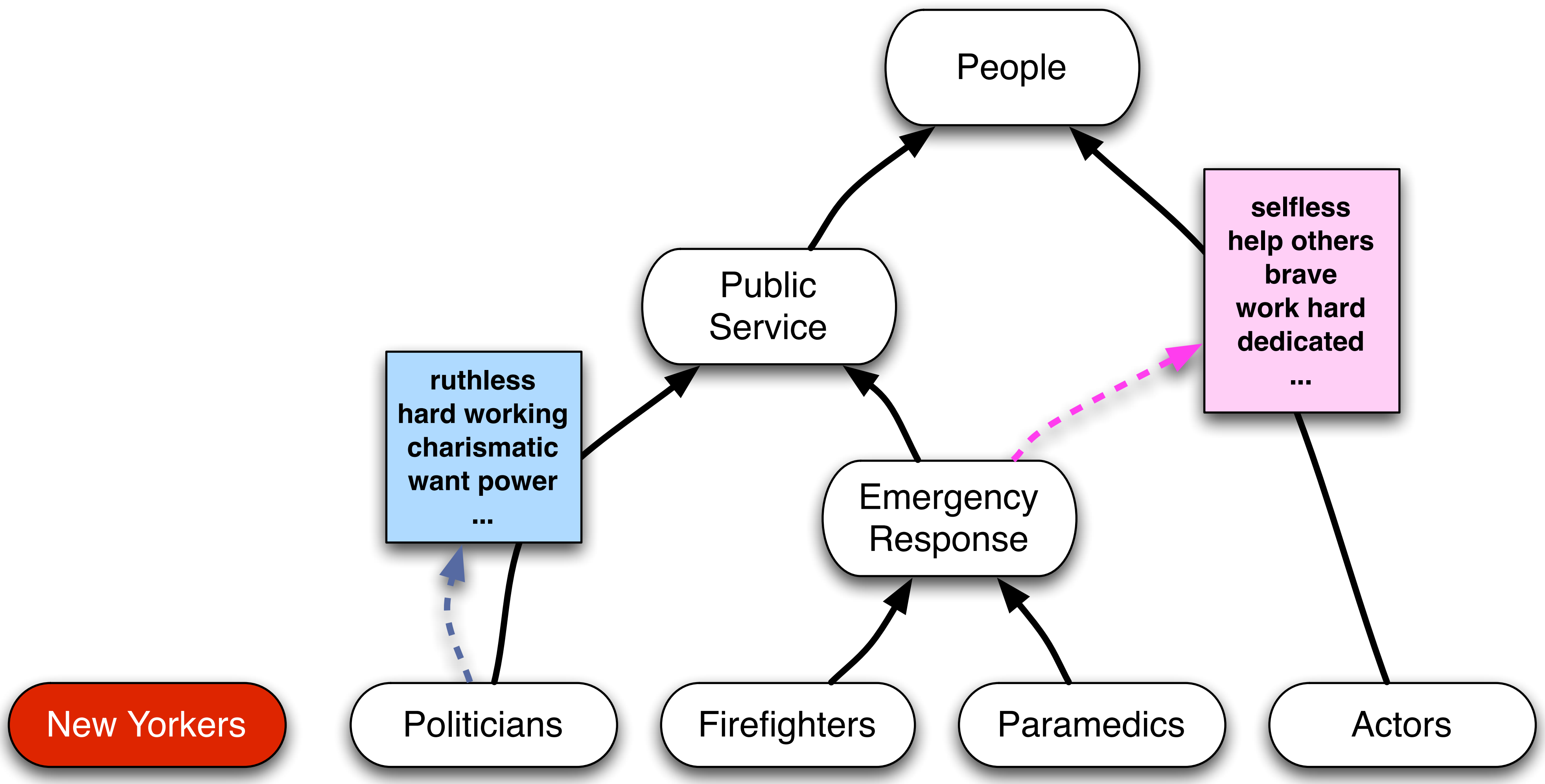

a reasonable hierarchical clustering of people

# A hypothesis



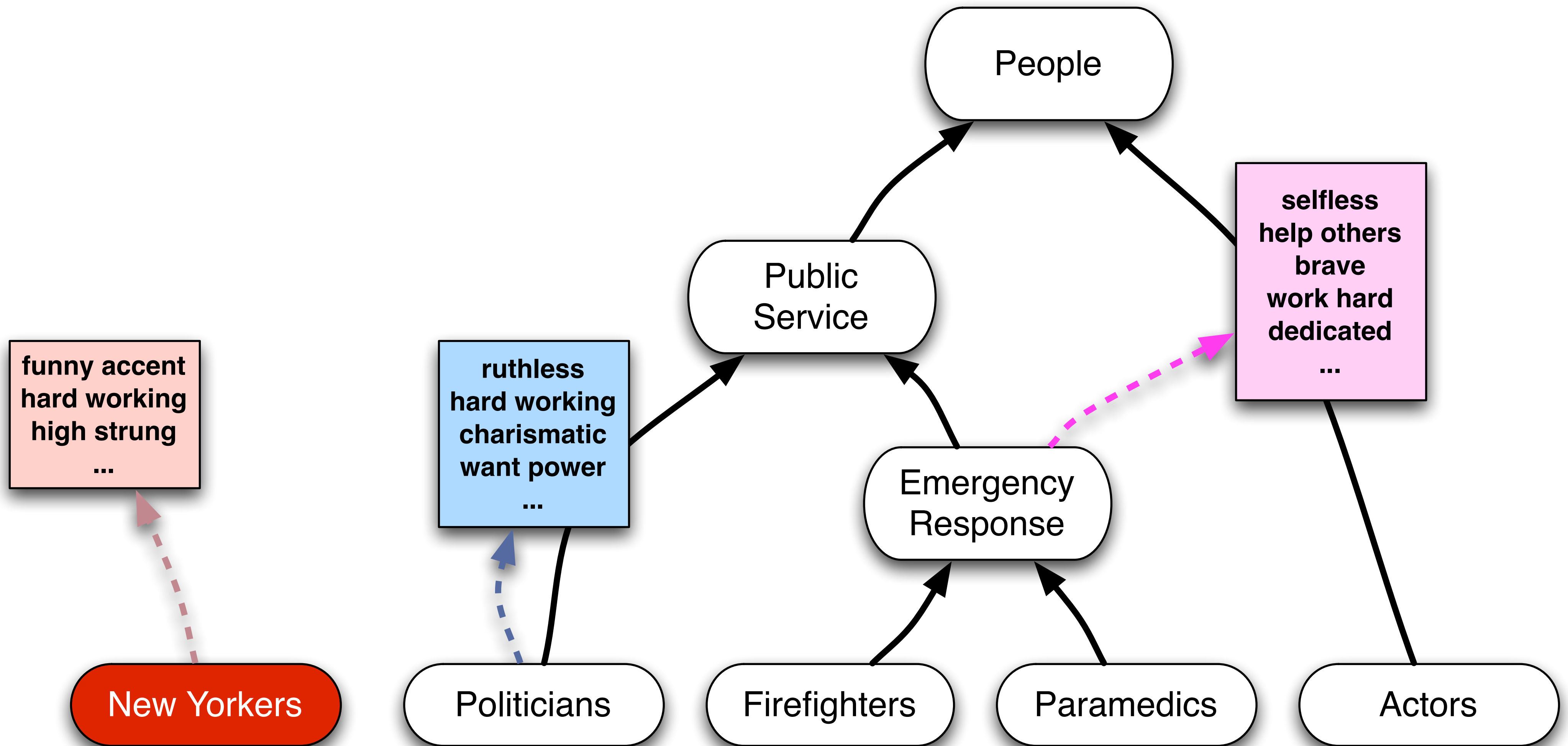

a reasonable hierarchical clustering of people

# A hypothesis



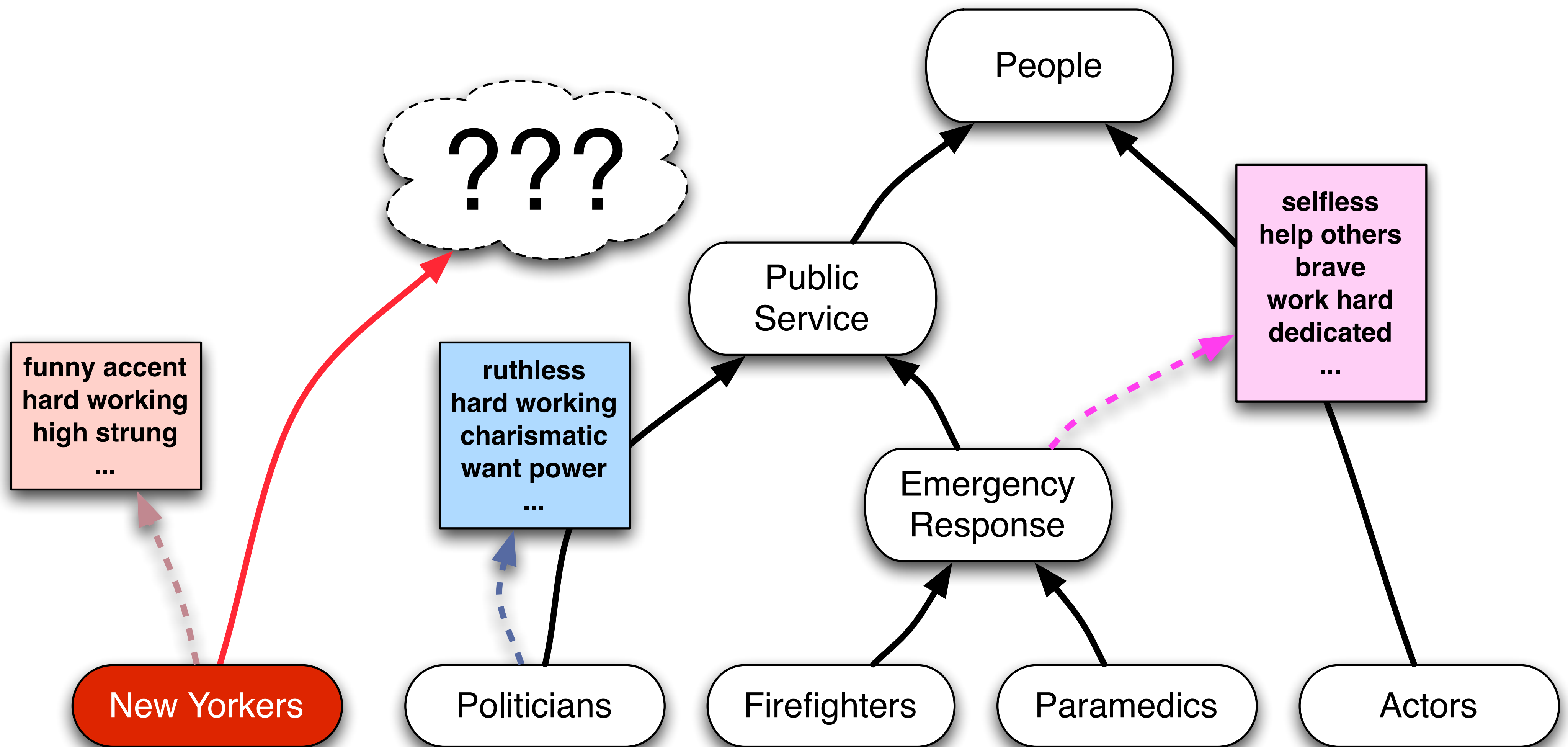

a reasonable hierarchical clustering of people

# A hypothesis




a reasonable hierarchical clustering of people

# A hypothesis




a reasonable hierarchical clustering of people

# A hypothesis

- “New Yorkers” belongs to a completely different, orthogonal categorization system, with a different set of salient features
- Same with “People born in 1961” and “Nobel Laureates”*
- Each categorization system controls what kinds of generalizations (e.g. inferences) are valid
- Want to account for Cross-cutting categorization

* btw all these examples come from Wikipedia

# A hypothesis

- Human conceptual organization drives lexical organization...
- i.e. these representational issues still exist at the lexical level
- In order to build effective lexical semantics models, we need to address human conceptual organization

# Empirically testable hypothesis

- Do word senses exhibit cross-cutting structure?
    - Xue, Chen and Palmer (2006): sense disambiguation requires vastly different features for different polysemous verbs in Chinese.
- What about verb arguments for selectional preference?
- Word relatedness?

# So, keep all of that in mind...

1) Simple models of concept organization can improve Web-based attribute extraction

2) Simple models of concept organization are predictive of the relatedness of words

3) What is it that these models are doing, exactly? (feature selection; hierarchical smoothing)

4) Generalizations based on cross-cutting categorization models

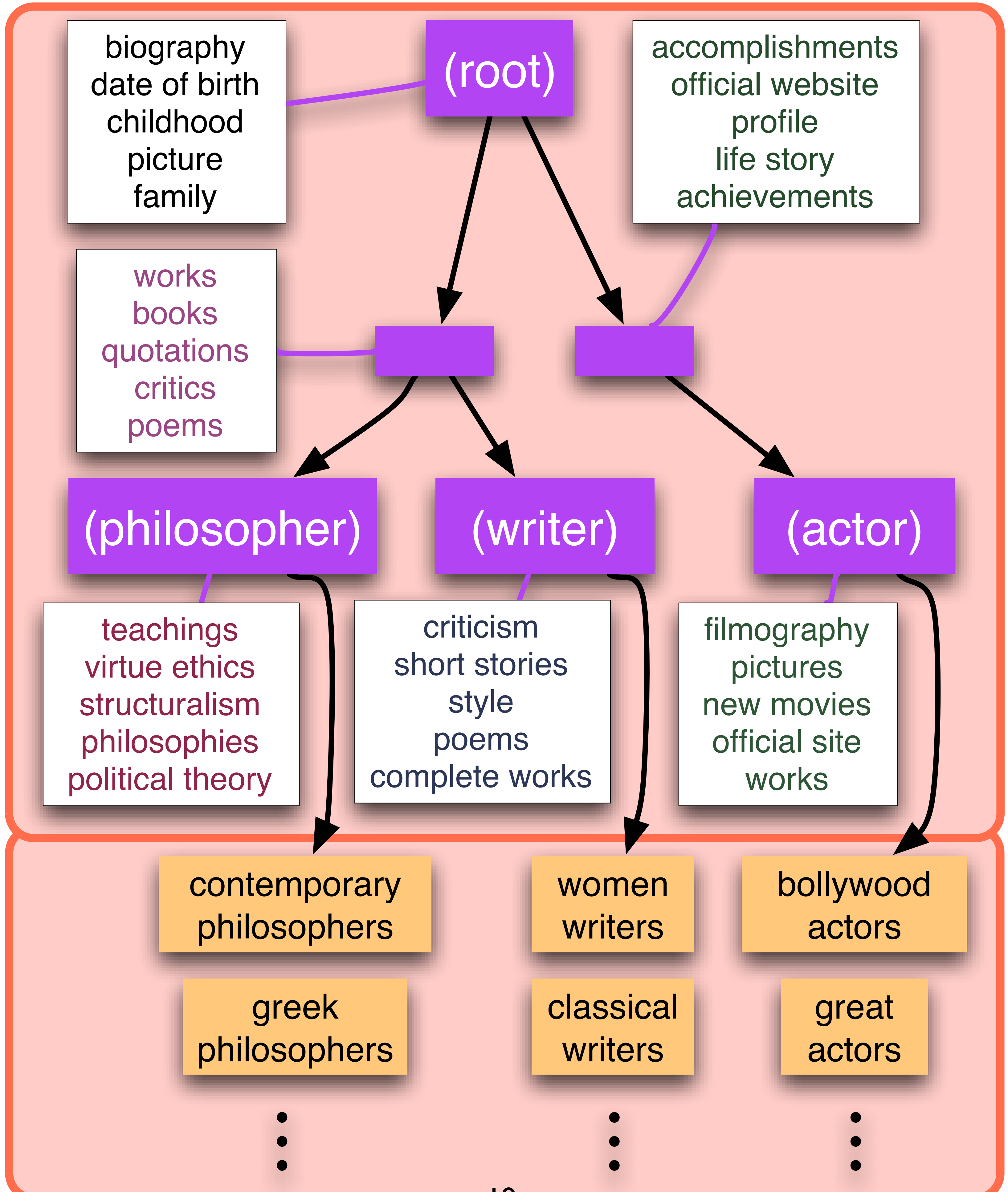


Shafto et al. (2006)

# Concept Organization

attribute-concept structure

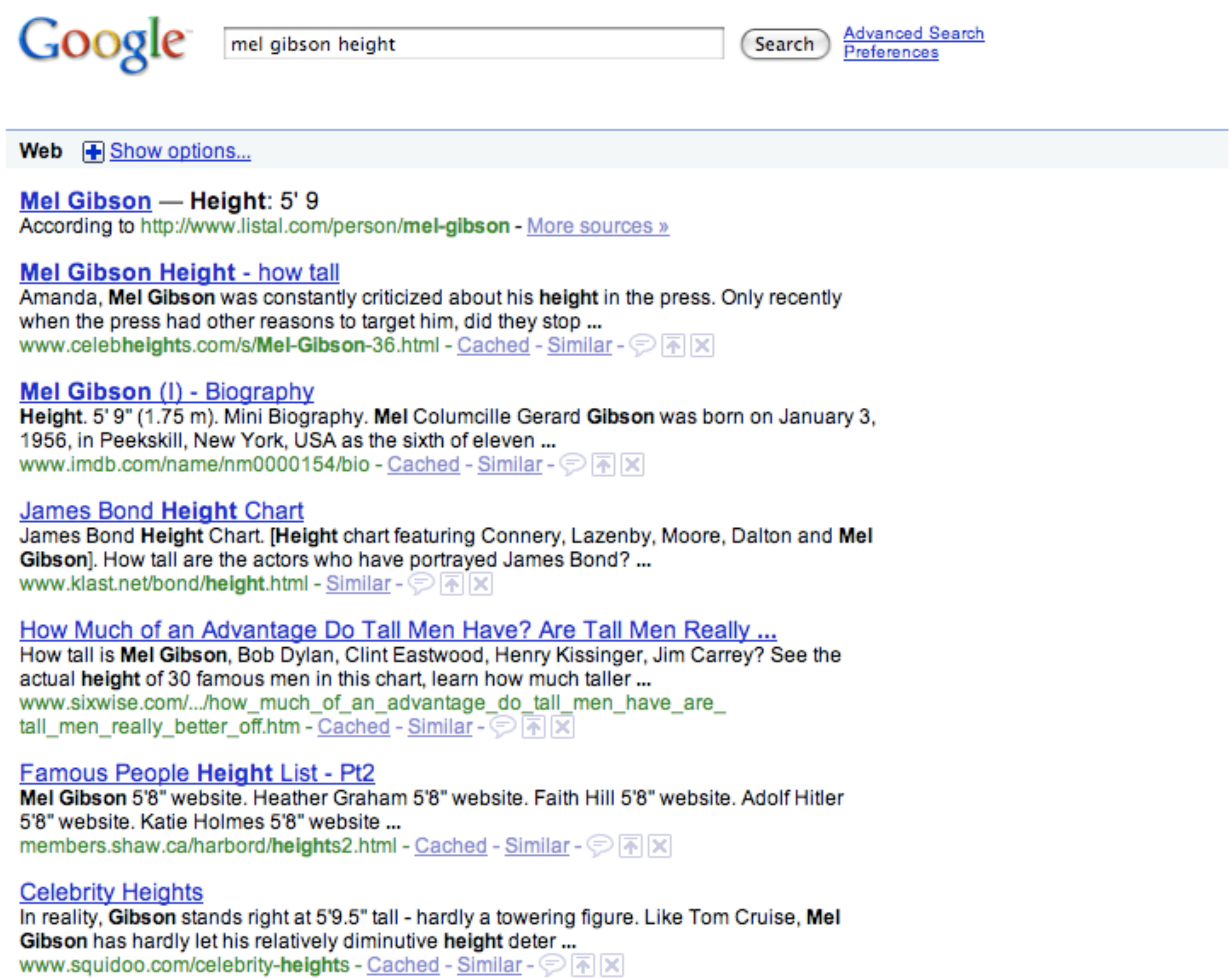
input data



# A little motivation

- Acquire facts for question answering
- IR, tail-query expansion
- Reduce noise in attribute/relation extraction
- Machine translation (e.g. anaphora resolution)

# Query logs



Google mel gibson height Search Advanced Search Preferences

Web Show options...

**Mel Gibson** — **Height**: 5' 9
According to http://www.listal.com/person/**mel-gibson** - More sources »

**Mel Gibson Height** - how tall
Amanda, **Mel Gibson** was constantly criticized about his **height** in the press. Only recently when the press had other reasons to target him, did they stop ...
www.celeb**heights**.com/s/**Mel-Gibson**-36.html - Cached - Similar

**Mel Gibson** (I) - Biography
**Height**. 5' 9" (1.75 m). Mini Biography. **Mel** Columcille Gerard **Gibson** was born on January 3, 1956, in Peekskill, New York, USA as the sixth of eleven ...
www.imdb.com/name/nm0000154/bio - Cached - Similar

James Bond **Height** Chart
James Bond **Height** Chart. [**Height** chart featuring Connery, Lazenby, Moore, Dalton and **Mel Gibson**]. How tall are the actors who have portrayed James Bond? ...
www.klast.net/bond/**height**.html - Similar

How Much of an Advantage Do Tall Men Have? Are Tall Men Really ...
How tall is **Mel Gibson**, Bob Dylan, Clint Eastwood, Henry Kissinger, Jim Carrey? See the actual **height** of 30 famous men in this chart, learn how much taller ...
www.sixwise.com/.../how_much_of_an_advantage_do_tall_men_have_are_tall_men_really_better_off.htm - Cached - Similar

Famous People **Height** List - Pt2
**Mel Gibson** 5'8" website. Heather Graham 5'8" website. Faith Hill 5'8" website. Adolf Hitler 5'8" website. Katie Holmes 5'8" website ...
members.shaw.ca/harbord/**heights**2.html - Cached - Similar

Celebrity Heights
In reality, **Gibson** stands right at 5'9.5" tall - hardly a towering figure. Like Tom Cruise, **Mel Gibson** has hardly let his relatively diminutive **height** deter ...
www.squidoo.com/celebrity-**heights** - Cached - Similar

# (class, instance) pairs

| | |
|---|---|
| **antineoplastic agents** | carmustine, dactinomycin, doxorubicin, fluorouracil, paclitaxel |
| **book publishers** | crown publishing, kluwer academic, prentice hall, puffin |
| **federal agencies** | catsa, dhs, dod, ex-im bank, tsis, iema, mema, nmfs, tdh, usdot |
| **mammals** | armadillo, elephant shrews, long-tailed weasel, river otter, wild goat |
| **scientific journals** | biometrika, european economic review, nature reviews genetics |
| **shipwrecks** | lusitania, mary celeste, bismarck, hms pandora, rms titanic |
| **social issues** | gender inequality, lack of education, substandard housing |
| **special diets** | kosher, lactose free, low-carb, peanut free, raw food, wheat-free |
| **turkish cities** | istanbul, kayseri, pergamum, balikesir, edirne, gaziantep, bursa |
| **turtles** | giant tortoise, painted turtle, red-eared slider, box turtle, flatback |
| **tyrants** | idi amin, justinian, emperor caligula, joseph stalin, genghis khan |
| **vulnerabilities** | denial of service, open relays, stolen passwords, spoofing |
| **writers** | bronte sisters, hemingway, kipling, proust, tasso, ungaretti, yeats |



(Van Durme & Paşca AAAI 2008)

# Noise in labeled attribute sets

| | |
|---|---|
| **antineoplastic agents** | mechanism of action, solubility, extravasation, contraindications |
| **book publishers** | adaptation, scientific name, adaptations, online dictionary, definition |
| **federal agencies** | castle, pay banding, locality pay, history, careers, secretary |
| **mammals** | digestive system, habitat, life cycle, respiratory system, reproduction |
| **scientific journals** | journal, impact factor, definition, archive, ranking, process, picture |
| **shipwrecks** | survivors, shipwreck, story, route, sinking, salvage, passenger list |
| **social issues** | health risks, cause and effect, definition, cartoons, meaning |
| **special diets** | definition, meaning, history, symptoms, low fat recipes, vitamins |
| **turkish cities** | population, history, climate, maps, weather, tourism, sightseeing |
| **turtles** | respiratory system, life cycle, sickness, habitat, drawing, predators |
| **tyrants** | autobiography, early life, childhood, mausoleum, bibliography |
| **vulnerabilities** | definition, history, list, different types, prevention, tutorial, statistics |
| **writers** | family crest, coat of arms, clan, family tree, bibliography, tartan |

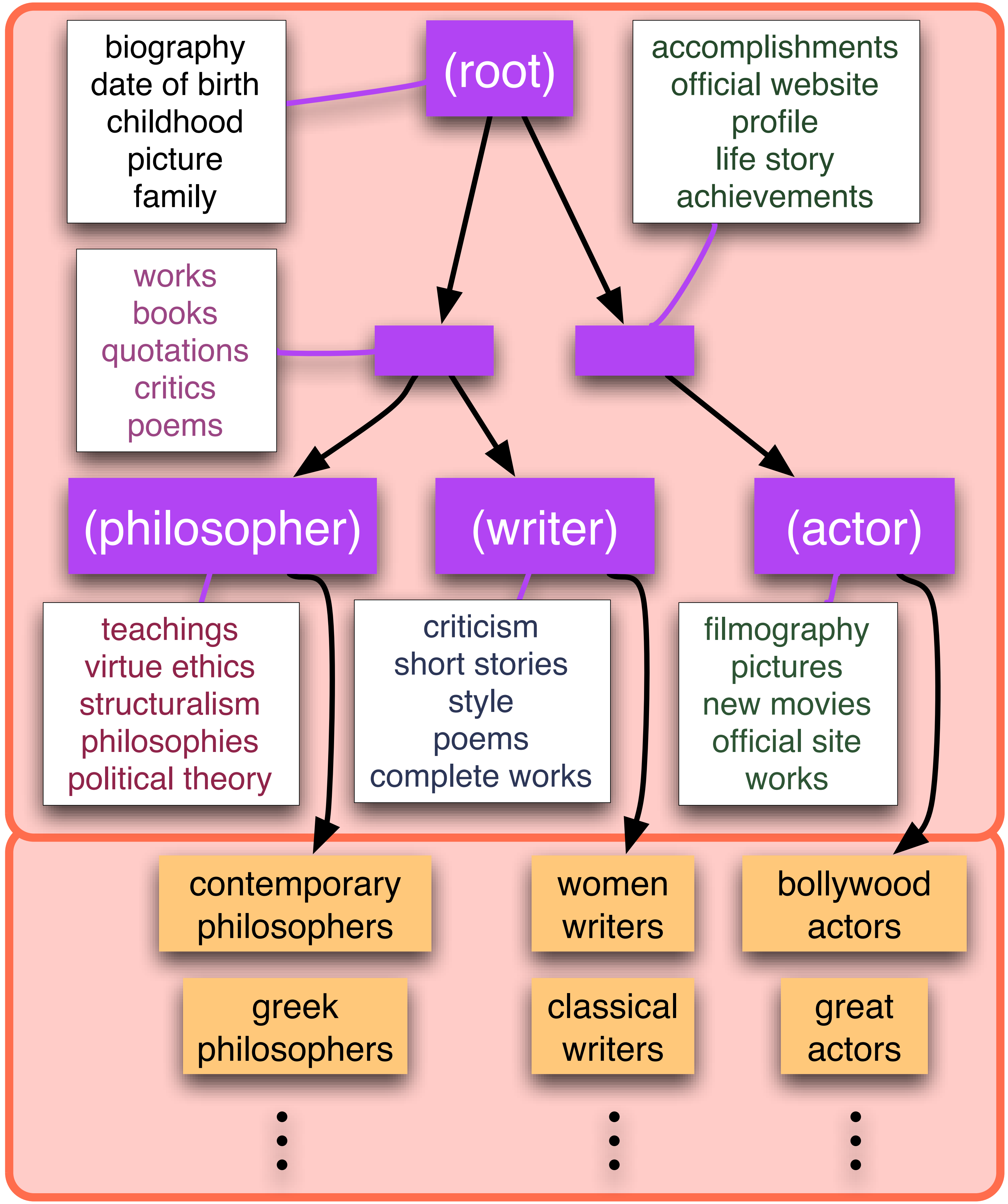
# Noise in labeled attribute sets

| | |
|---|---|
| antineoplastic agents | **mechanism of action, solubility, extravasation, contraindications** |
| book publishers | adaptation, scientific name, adaptations, online dictionary, definition |
| federal agencies | castle, pay banding, locality pay, history, careers, secretary |
| mammals | **digestive system, habitat, life cycle, respiratory system, reproduction** |
| scientific journals | journal, impact factor, definition, archive, ranking, process, picture |
| shipwrecks | survivors, shipwreck, story, route, sinking, salvage, passenger list |
| social issues | health risks, cause and effect, definition, cartoons, meaning |
| special diets | definition, meaning, history, symptoms, low fat recipes, vitamins |
| turkish cities | **population, history, climate, maps, weather, tourism, sightseeing** |
| turtles | **respiratory system, life cycle, sickness, habitat, drawing, predators** |
| tyrants | autobiography, early life, childhood, mausoleum, bibliography |
| vulnerabilities | definition, history, list, different types, prevention, tutorial, statistics |
| writers | family crest, coat of arms, clan, family tree, bibliography, tartan |

# Noise in labeled attribute sets

| | |
|---|---|
| antineoplastic agents | mechanism of action, solubility, extravasation, contraindications |
| book publishers | **adaptation, scientific name, adaptations, online dictionary, definition** |
| federal agencies | **castle, pay banding, locality pay, history, careers, secretary** |
| mammals | digestive system, habitat, life cycle, respiratory system, reproduction |
| scientific journals | journal, impact factor, definition, archive, ranking, process, picture |
| shipwrecks | survivors, shipwreck, story, route, sinking, salvage, passenger list |
| social issues | health risks, cause and effect, definition, cartoons, meaning |
| special diets | **definition, meaning, history, symptoms, low fat recipes, vitamins** |
| turkish cities | population, history, climate, maps, weather, tourism, sightseeing |
| turtles | respiratory system, life cycle, sickness, habitat, drawing, predators |
| tyrants | autobiography, early life, childhood, mausoleum, bibliography |
| vulnerabilities | **definition, history, list, different types, prevention, tutorial, statistics** |
| writers | **family crest, coat of arms, clan, family tree, bibliography, tartan** |

attribute-concept structure




input data

# Topic Models

## Latent Dirichlet Allocation

$$
\begin{array}{rclll}
\boldsymbol{\theta}_d|\boldsymbol{\alpha} & \sim & \mathrm{Dirichlet}(\boldsymbol{\alpha}), & d \in D, & \text{(topic proportions)} \\
\boldsymbol{\phi}_t|\boldsymbol{\beta} & \sim & \mathrm{Dirichlet}(\boldsymbol{\beta}), & t \in T, & \text{(topics)} \\
z_{id}|\boldsymbol{\theta}_d & \sim & \mathrm{Mult}(\boldsymbol{\theta}_d), & i \in |\mathbf{w}_d|, & \text{(topic indicators)} \\
w_{id}|\boldsymbol{\phi}_{z_{id}} & \sim & \mathrm{Mult}(\boldsymbol{\phi}_{z_{id}}), & i \in |\mathbf{w}_d|, & \text{(words)}
\end{array}
$$

| $\phi_1$ | $\phi_2$ | $\phi_3$ |
|---|---|---|
| government | wrote | finance |
| minister | said | economists |
| state | responding | spending |
| federal | editor | budget |

$\sim \mathrm{Dir}(\beta)$

$\boldsymbol{\phi} =$ 

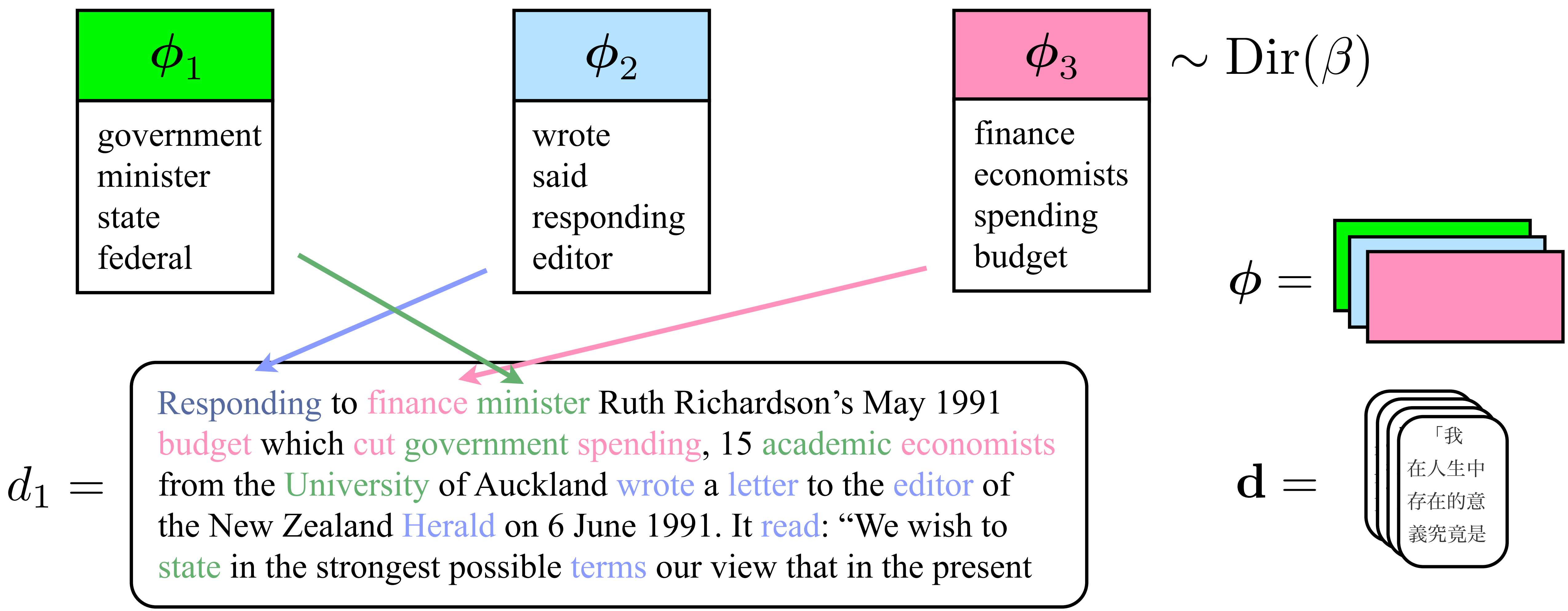
$d_1 =$

$\mathbf{d} =$ 

# Topic Models

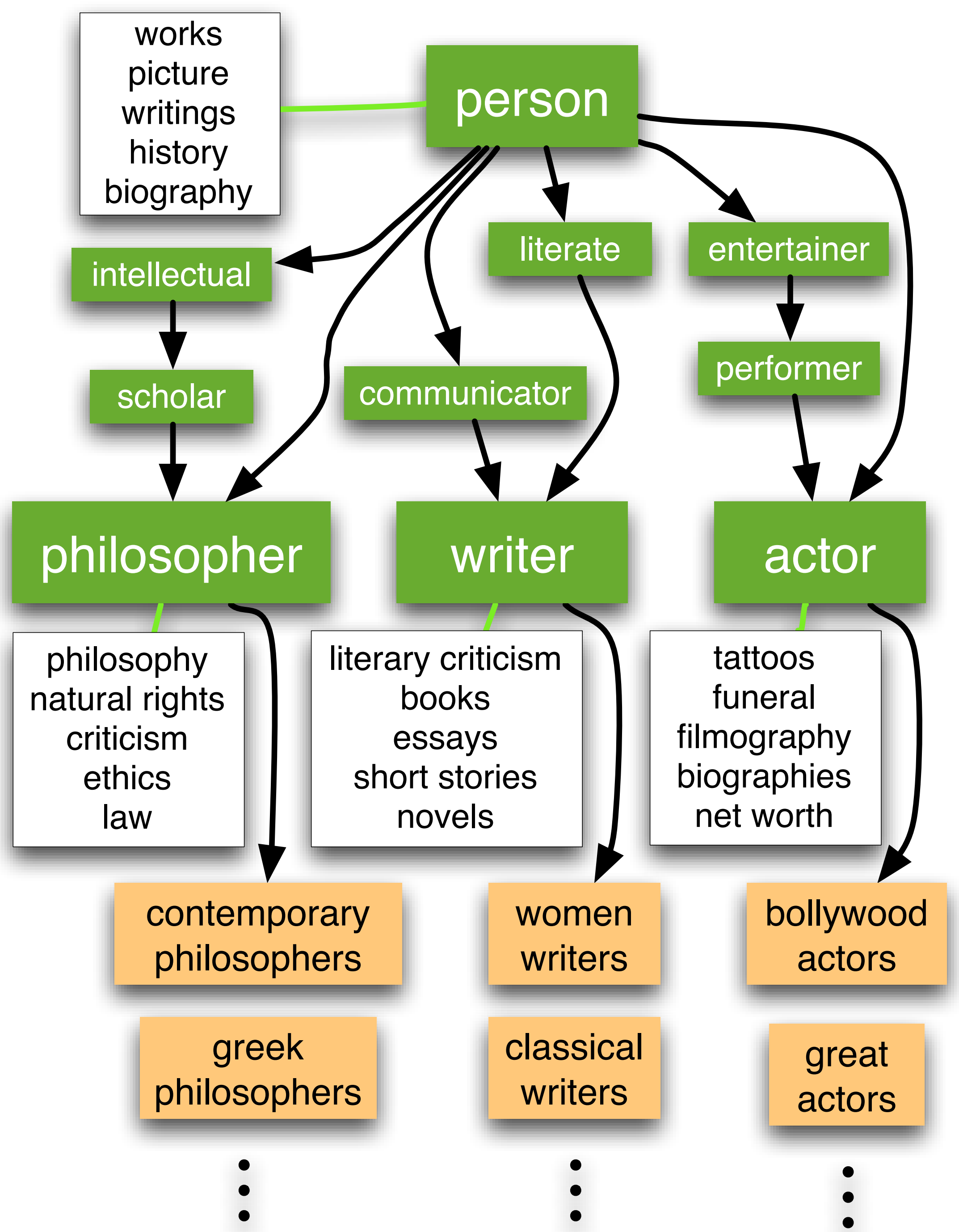
## Latent Dirichlet Allocation

$$
\begin{array}{rcllll}
\boldsymbol{\theta}_d | \boldsymbol{\alpha} & \sim & \text{Dirichlet}(\boldsymbol{\alpha}), & d \in D, & \text{(topic proportions)} \\
\boldsymbol{\phi}_t | \boldsymbol{\beta} & \sim & \text{Dirichlet}(\boldsymbol{\beta}), & t \in T, & \text{(topics)} \\
z_{id} | \boldsymbol{\theta}_d & \sim & \text{Mult}(\boldsymbol{\theta}_d), & i \in |\mathbf{w}_d|, & \text{(topic indicators)} \\
w_{id} | \boldsymbol{\phi}_{z_{id}} & \sim & \text{Mult}(\boldsymbol{\phi}_{z_{id}}), & i \in |\mathbf{w}_d|, & \text{(words)}
\end{array}
$$

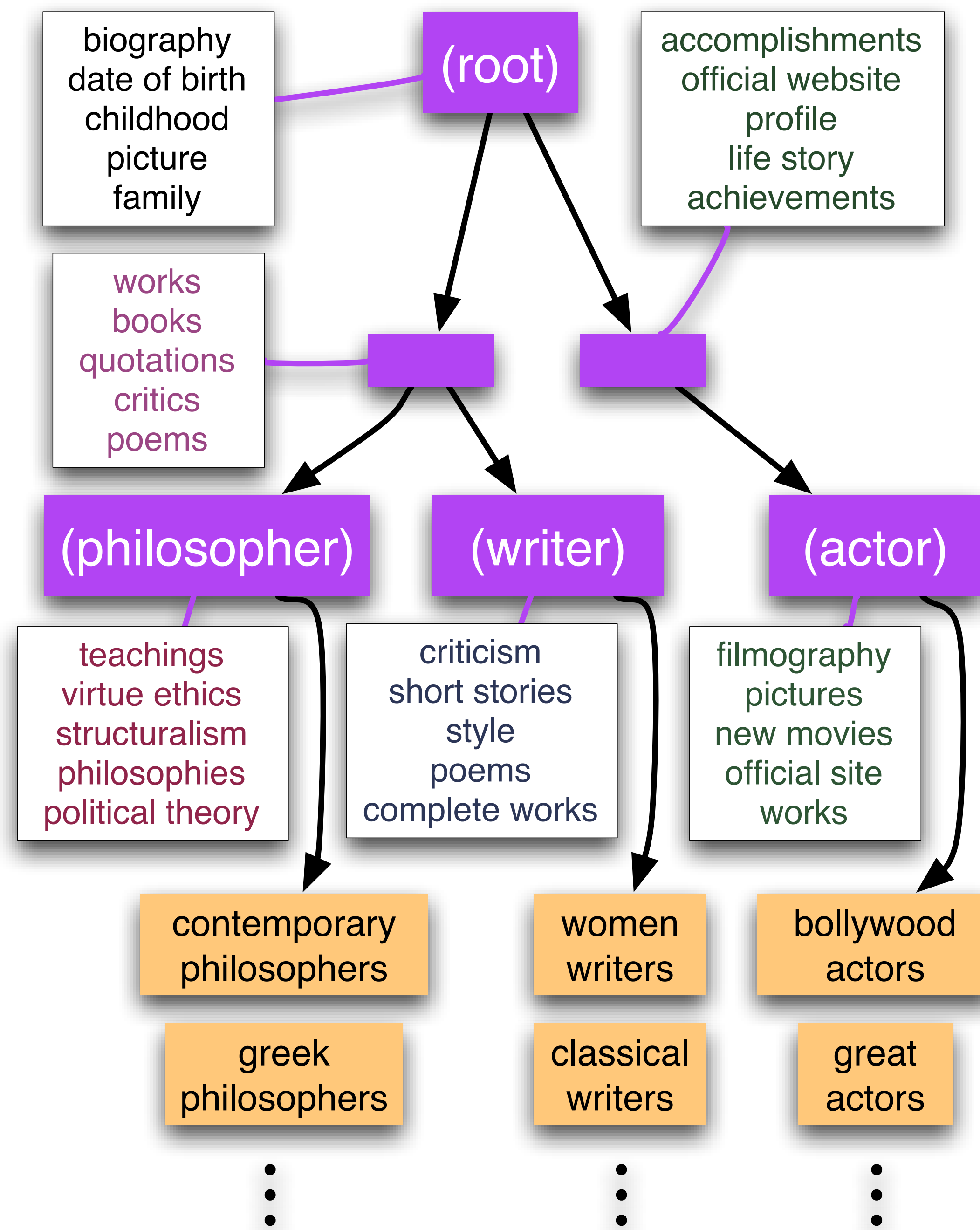


# Two kinds of (non cross-cutting) structure

## Fixed (LDA+WordNet)




## Learned (hLDA)

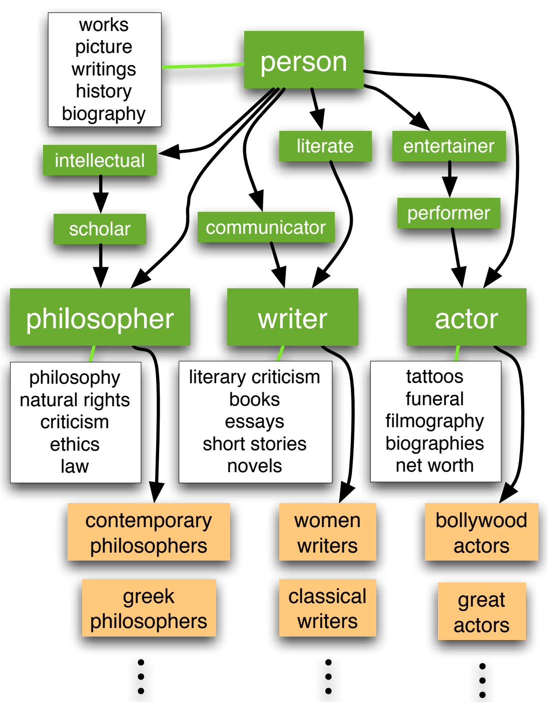





2010年7月17日土曜日

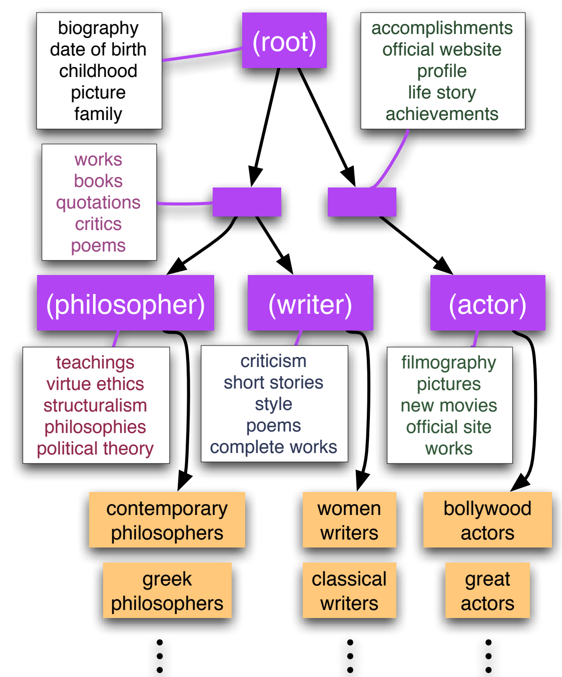
# Advantages / disadvantages

## fixed (LDA+wordnet)



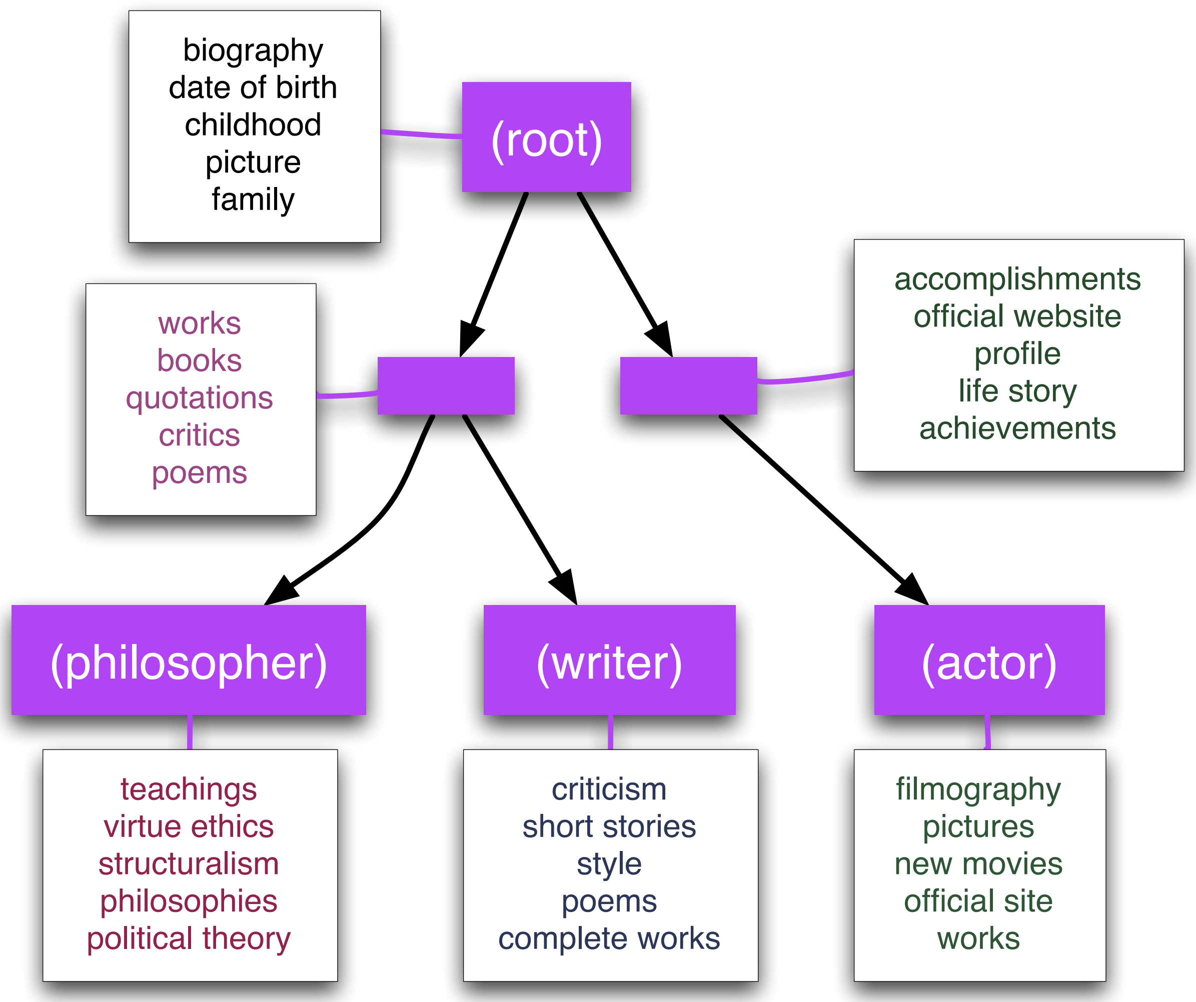

- Worse precision
- Human-understandable intermediate concepts

## learned (hLDA)




- Higher precision as a smoother
- Hard to evaluate intermediate concepts

# Generative process

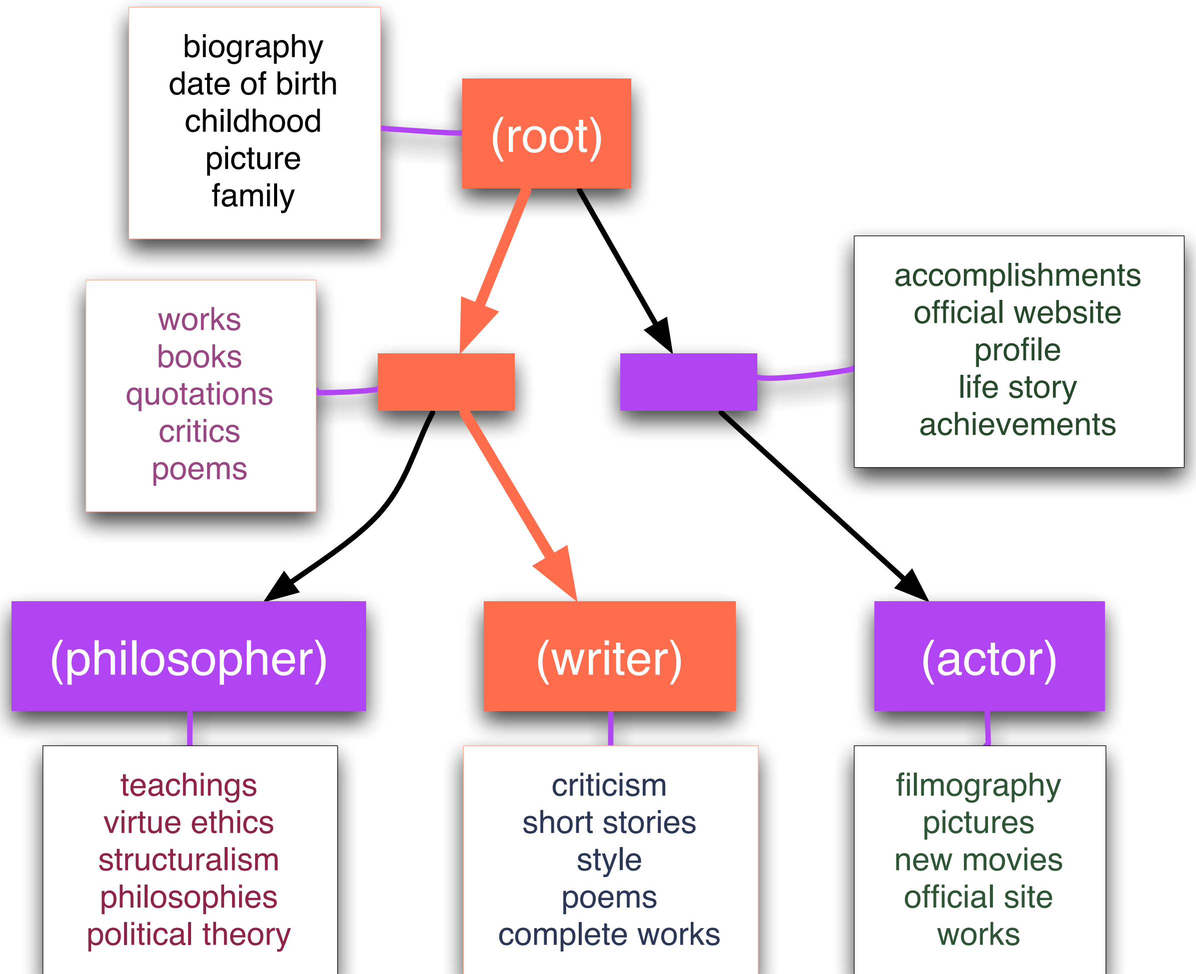



For each "document" $d$ :
- Choose a path from the DAG $\mathbf{c}_d$
- Choose a multinomial over levels, $\theta_d$

For each "word" $w_{i,d}$ :
- Choose a level $z_{i,d}$
- Choose a word from that node's distribution $p(\mathbf{w}_d|\mathbf{c},\mathbf{z})$

document = attributes for class X
word = attribute

# Generative process




short story writer:

For each "document" $d$ :

- Choose a path from the DAG $\mathbf{c}_d$ [either uniform over paths or from nCRP]
- Choose a multinomial over levels, $\theta_d$

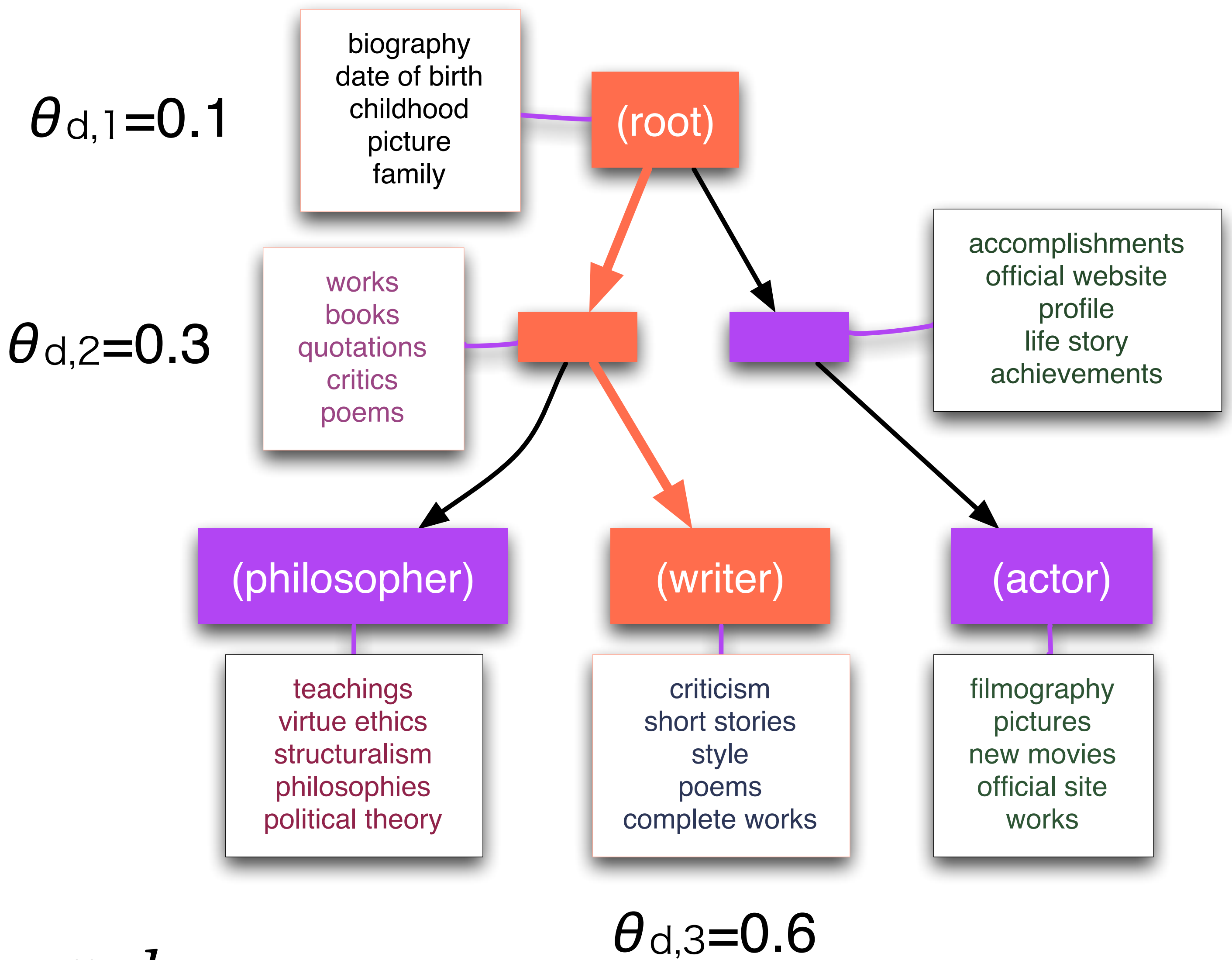
For each "word" $w_{i,d}$ :

- Choose a level $z_{i,d}$
- Choose a word from that node's distribution $p(\mathbf{w}_d|\mathbf{c},\mathbf{z})$

document = attributes for class X
word = attribute

# Generative process

short story writer:




For each "document" $d$ :

- Choose a path from the DAG $\mathbf{c}_d$
- Choose a multinomial over levels, $\theta_d$

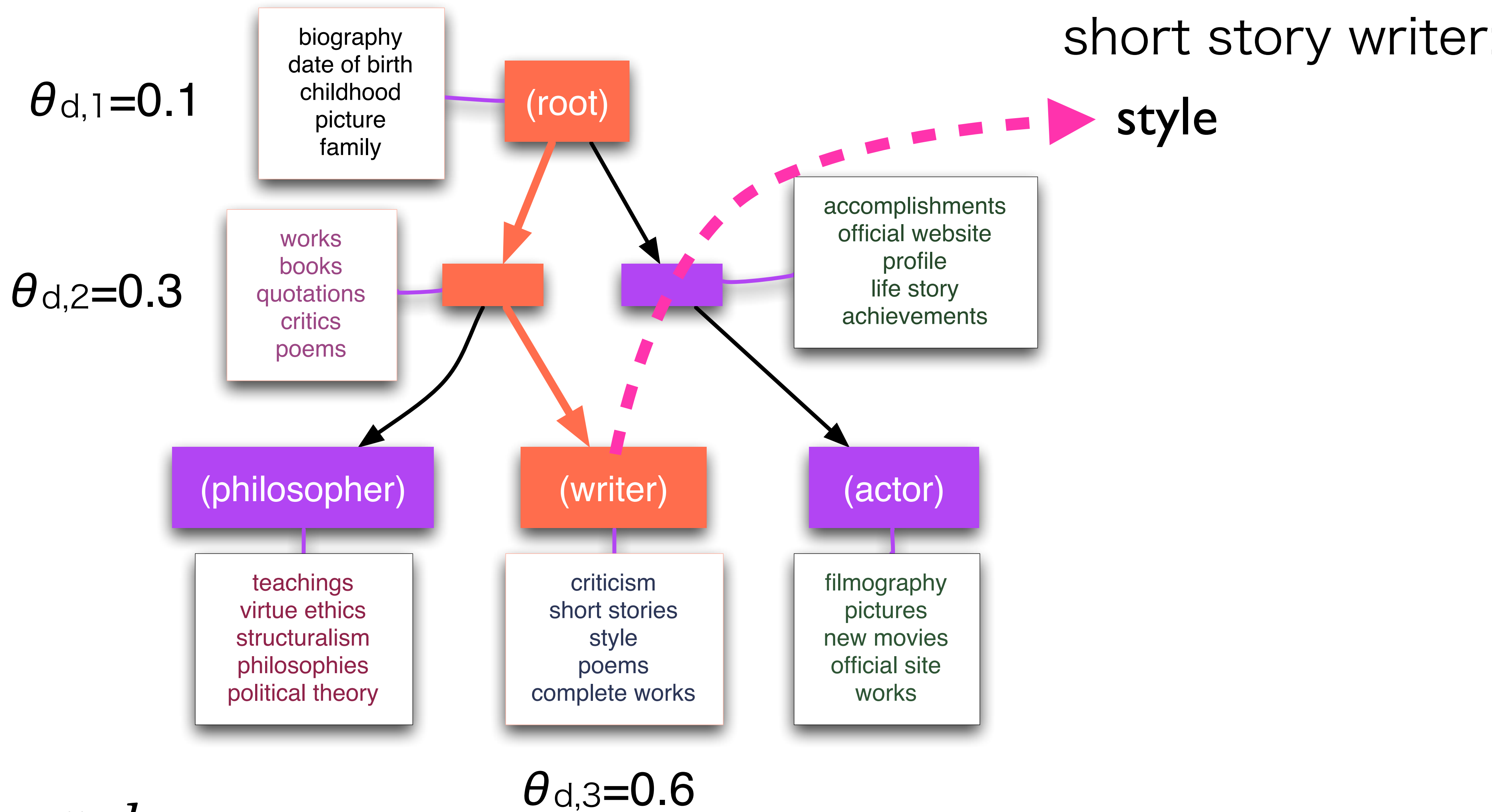
For each "word" $w_{i,d}$ :

- Choose a level $z_{i,d}$
- Choose a word from that node's distribution $p(\mathbf{w}_d|\mathbf{c},\mathbf{z})$

document = attributes for class X

word = attribute

# Generative process




For each "document" $d$ :

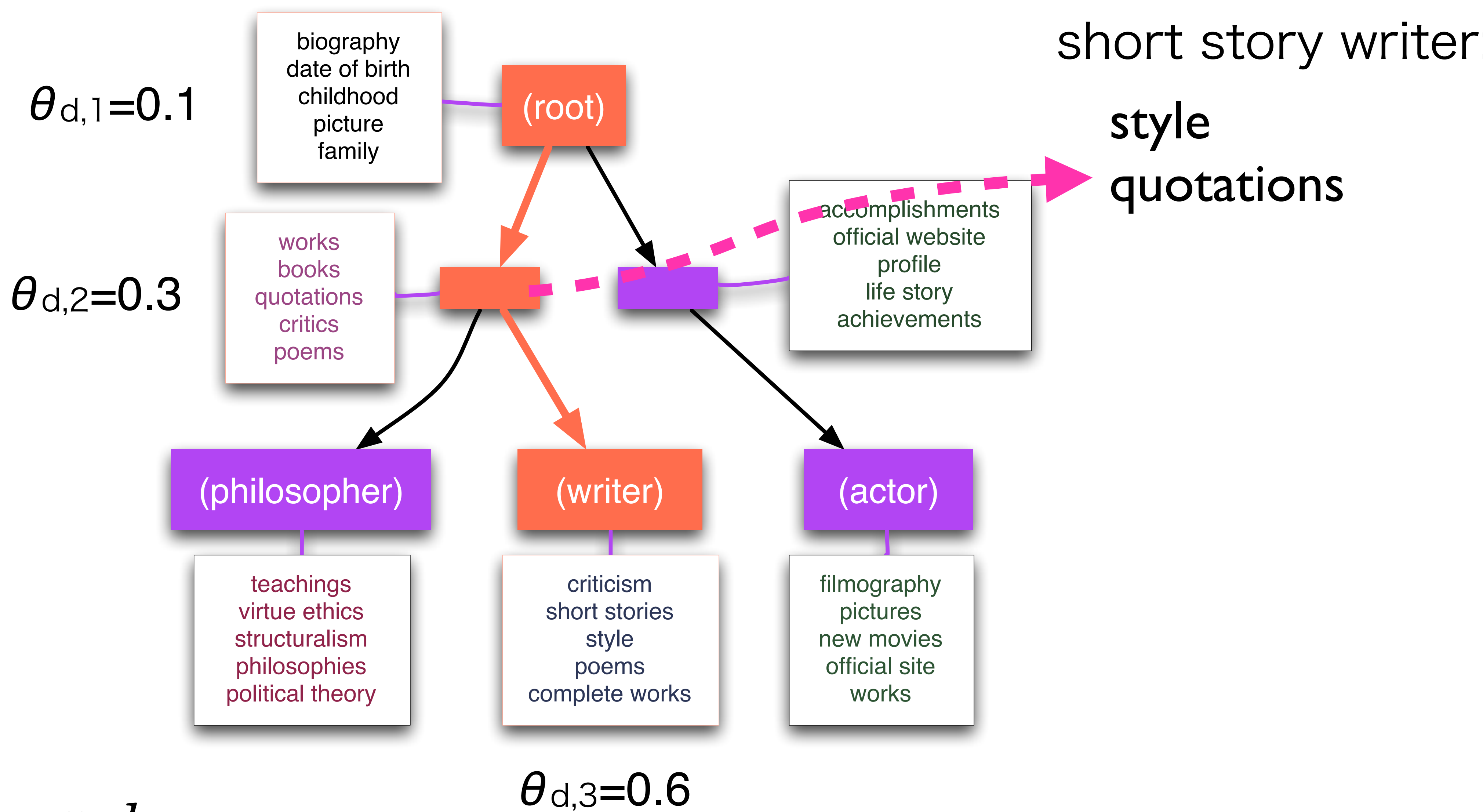
- Choose a path from the DAG $\mathbf{c}_d$
- Choose a multinomial over levels, $\theta_d$

For each "word" $w_{i,d}$ :

- Choose a level $z_{i,d}$
- Choose a word from that node's distribution $p(\mathbf{w}_d|\mathbf{c},\mathbf{z})$

document = attributes for class X
word = attribute

# Generative process




For each "document" $d$ :

- Choose a path from the DAG $\mathbf{c}_d$
- Choose a multinomial over levels, $\theta_d$

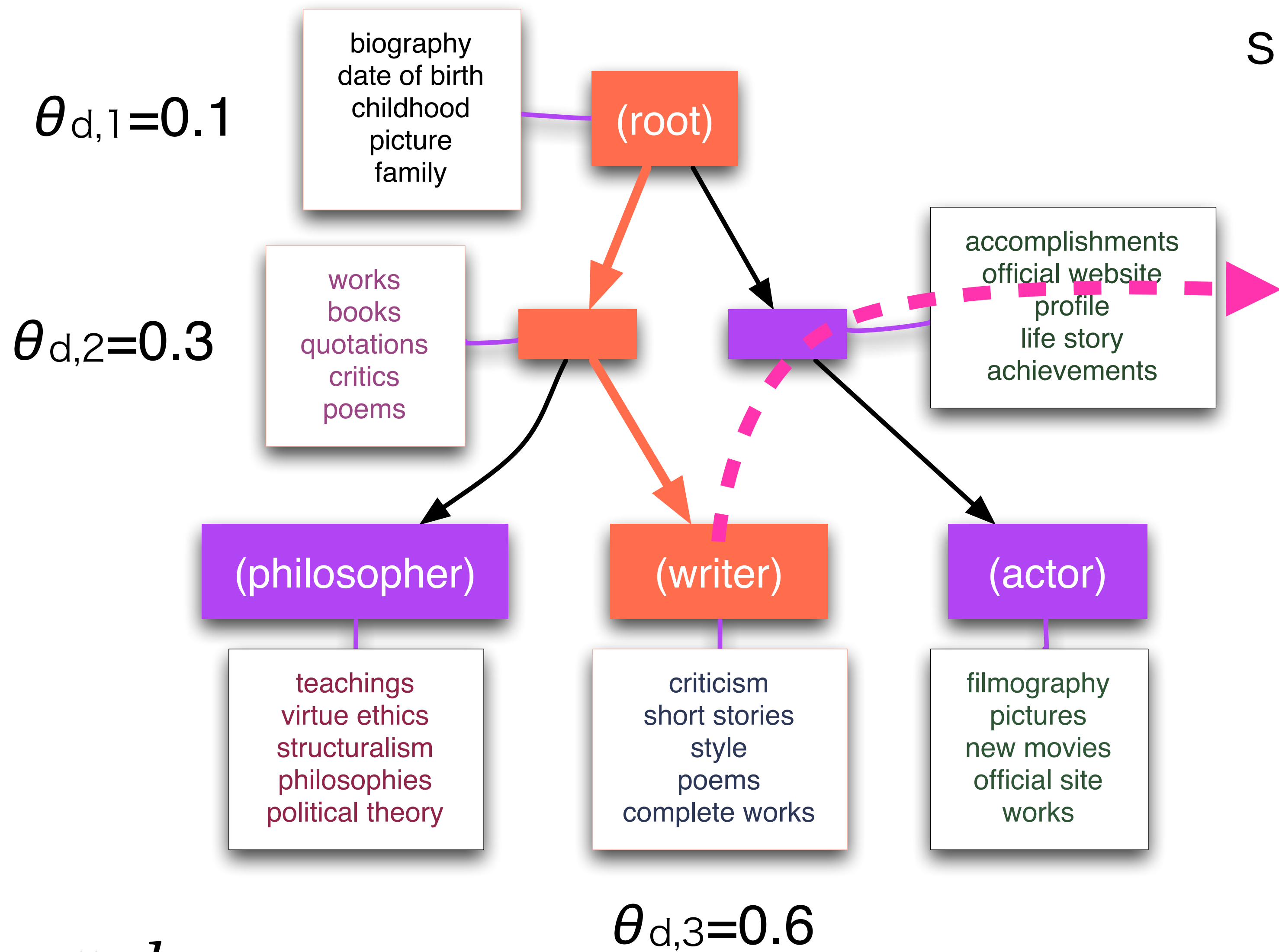
For each "word" $w_{i,d}$ :

- Choose a level $z_{i,d}$
- Choose a word from that node's distribution $p(\mathbf{w}_d | \mathbf{c}, \mathbf{z})$

document = attributes for class X

word = attribute

# Generative process




short story writer:

**style**
**quotations**
**short stories**

For each "document" $d$ :

- Choose a path from the DAG $\mathbf{c}_d$
- Choose a multinomial over levels, $\theta_d$

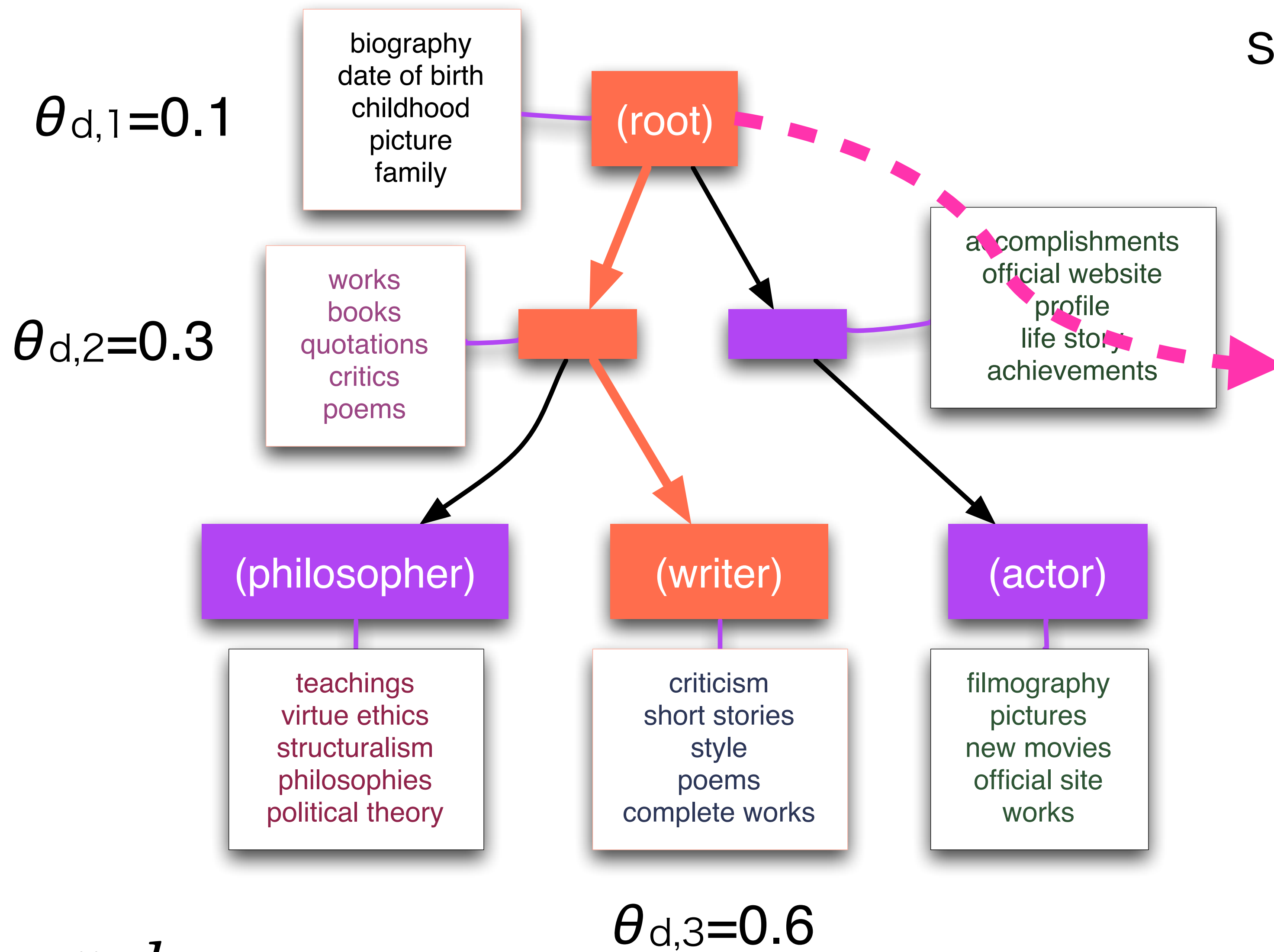
For each "word" $w_{i,d}$ :

- Choose a level $z_{i,d}$
- Choose a word from that node's distribution $p(\mathbf{w}_d|\mathbf{c}, \mathbf{z})$

document = attributes for class X

word = attribute

# Generative process




short story writer:

style
quotations
short stories
biography

For each "document" $d$ :

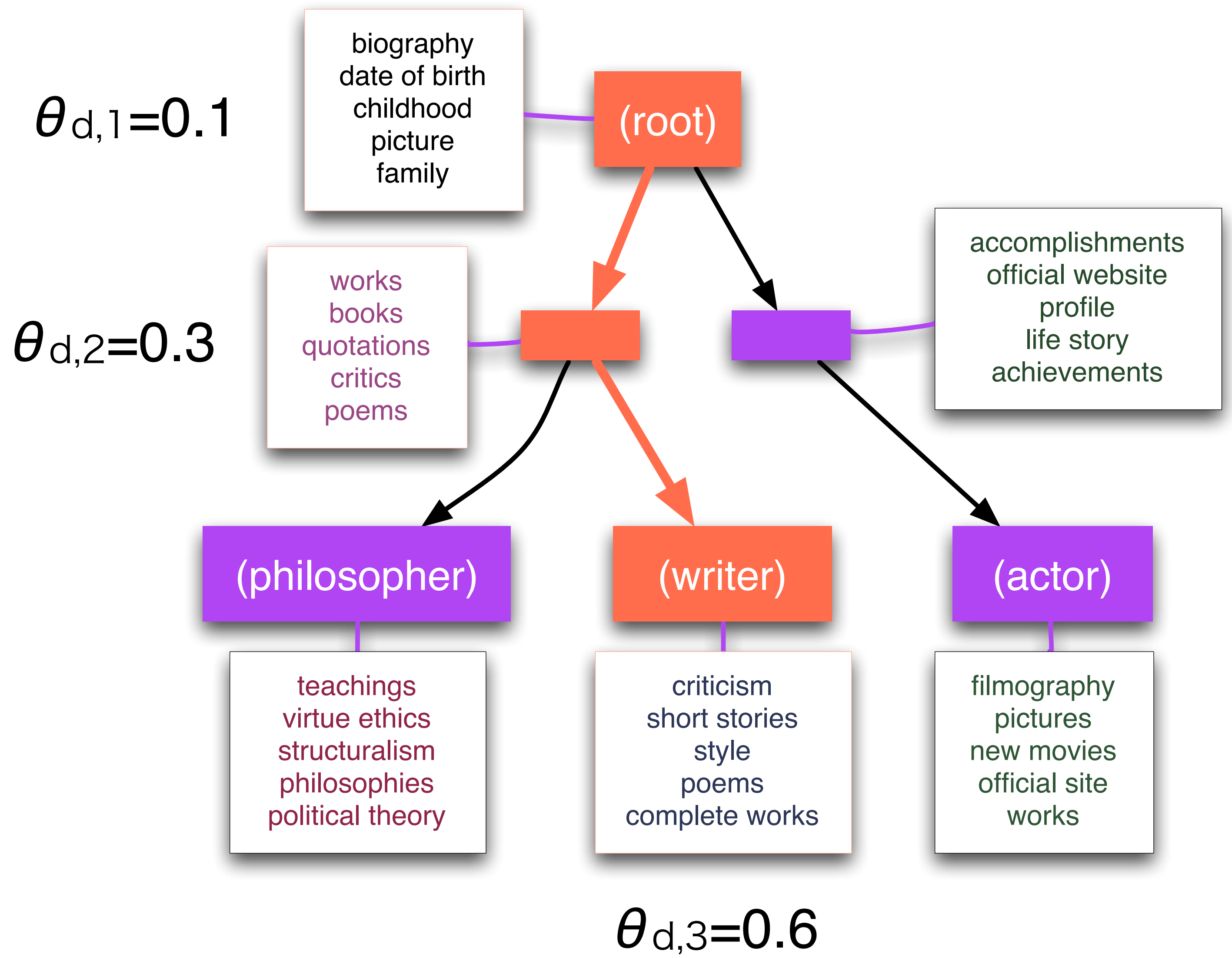
- Choose a path from the DAG $\mathbf{c}_d$
- Choose a multinomial over levels, $\theta_d$

For each "word" $w_{i,d}$ :

- Choose a level $z_{i,d}$
- Choose a word from that node's distribution $p(\mathbf{w}_d|\mathbf{c},\mathbf{z})$

document = attributes for class X
word = attribute

# Generative process




short story writer:

style
quotations
short stories
biography

- Why hLDA:
  - Semantically distinct attribute distributions
  - Extensible to more complex structure

# Empirical evaluation

(1) Attribute re-ranking / noise-filtering / smoothing precision

$$p_{\text{lda}}(w|\mathbf{w}_d) := \sum_c p(w|c,\eta)p(c|\mathbf{w}_d,\alpha)$$

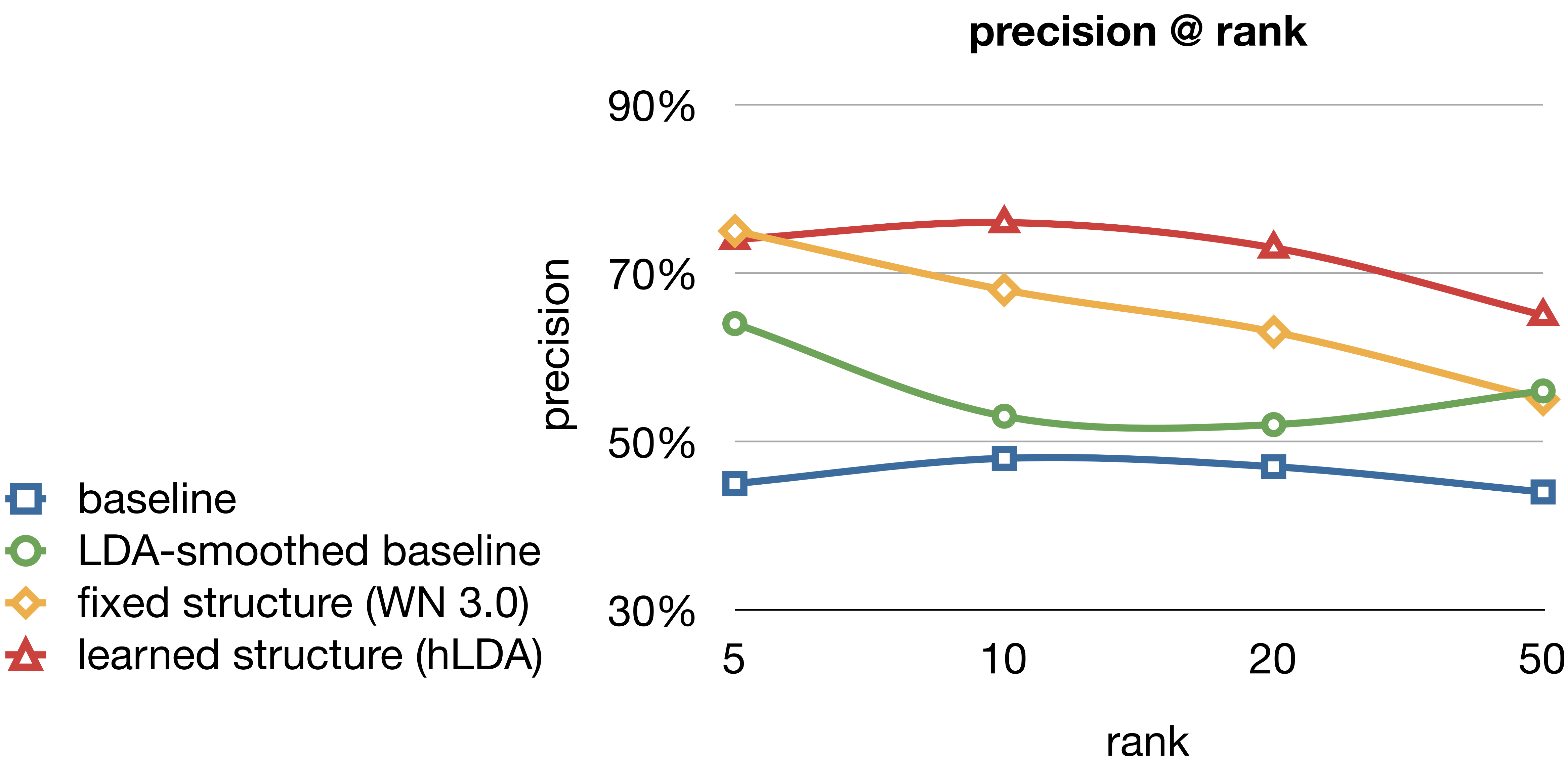
$$p(w|\mathbf{w}_d) = p_{\text{lda}}(w|\mathbf{w}_d)p_{\text{base}}(w|\mathbf{w}_d) \qquad p_{\text{base}}(w|\mathbf{w}_d) \stackrel{\text{def}}{=} \theta^{r(w,\mathbf{w}_d)}$$

(2) Concept assignment precision (determining the right degree of specificity)

baseline: propagate attributes up the tree

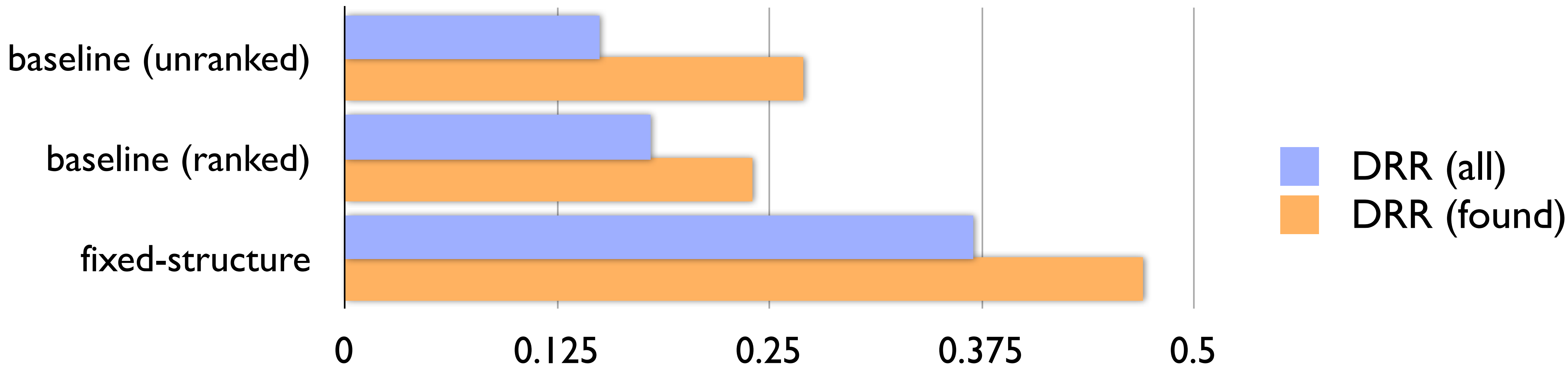
(Paşca AAAI 2008)

# Results: re-ranked attribute precision



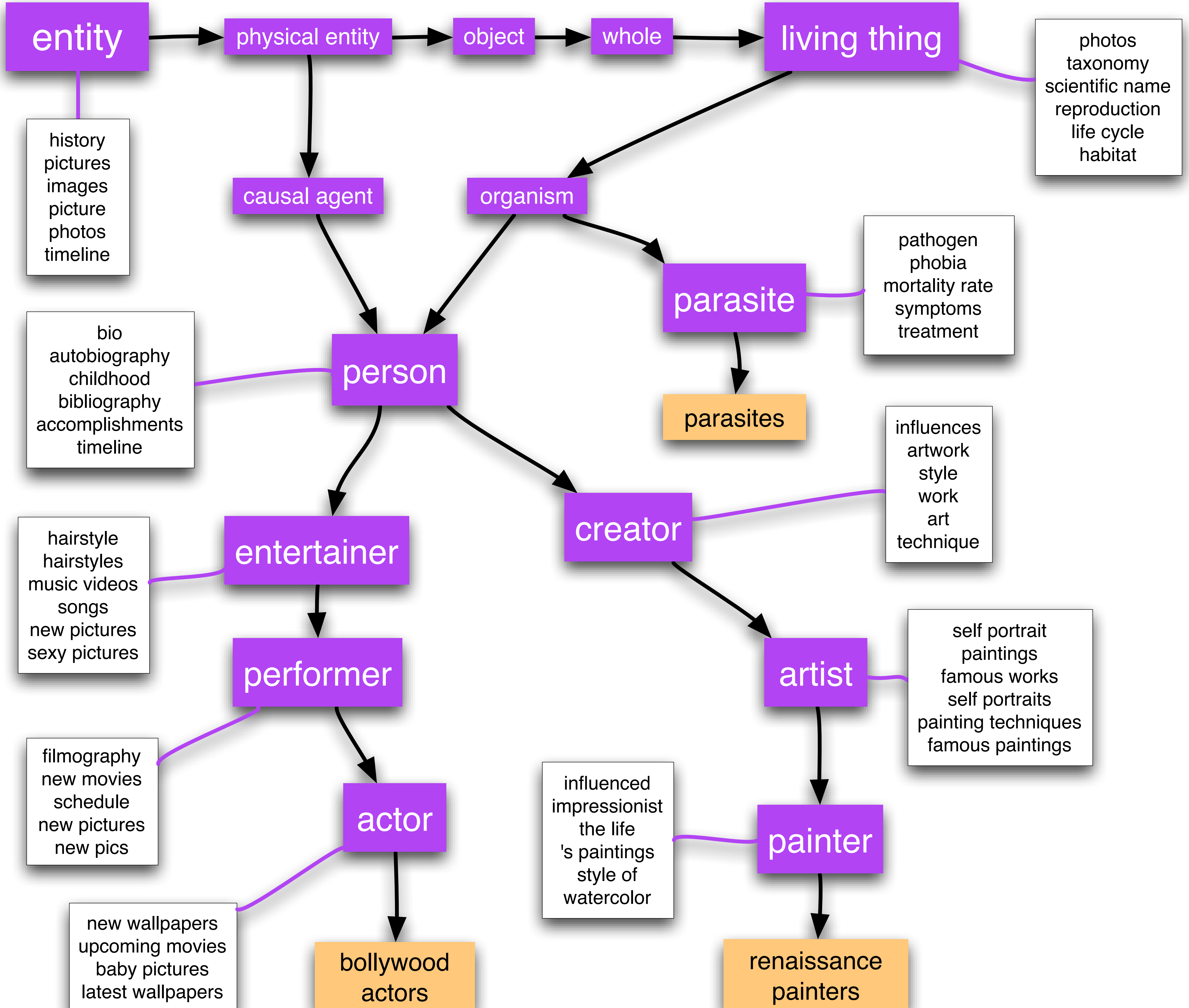

- Precision scores from human raters
- hLDA smoothing significantly improves precision over ranked baseline

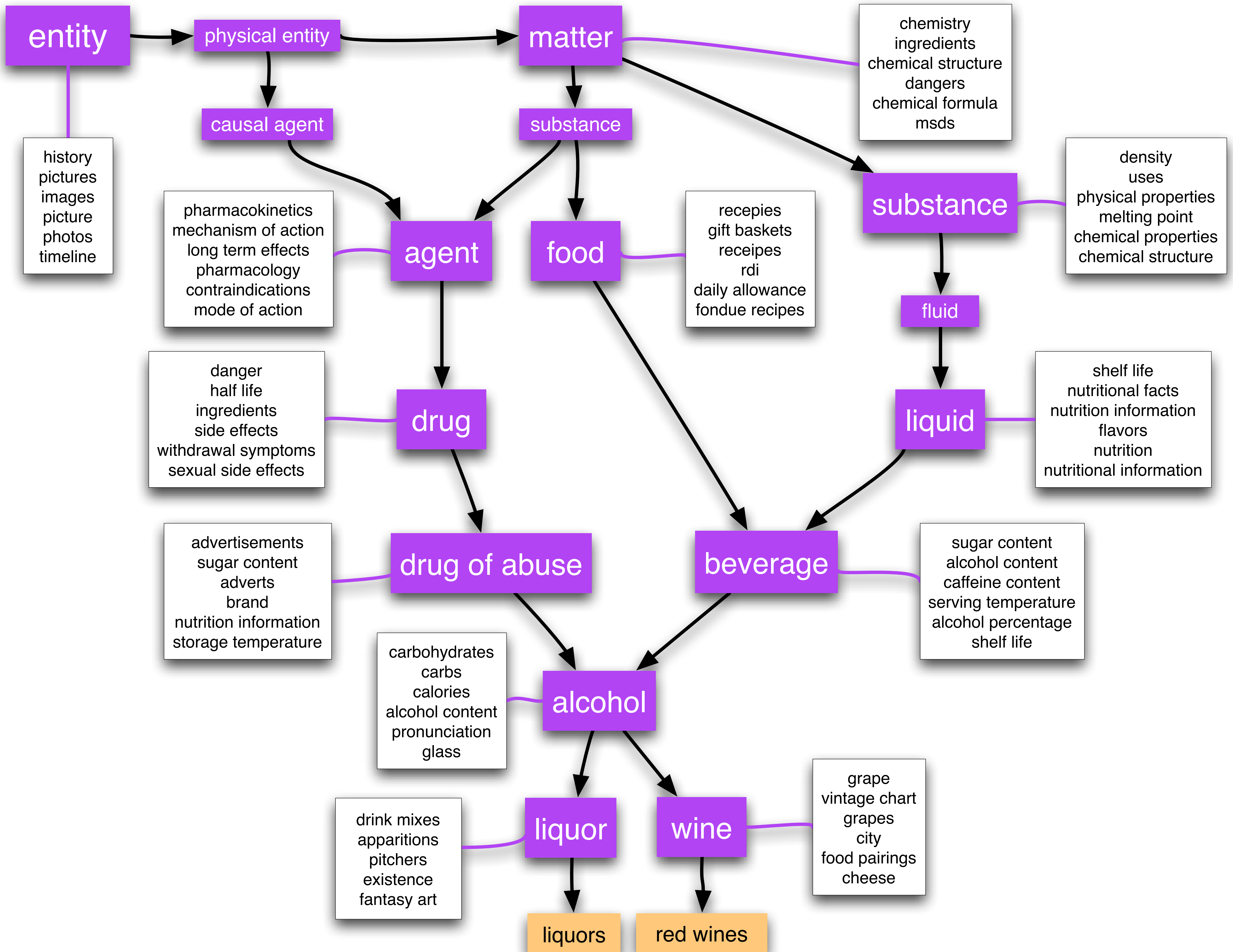
# Results: concept assignment precision

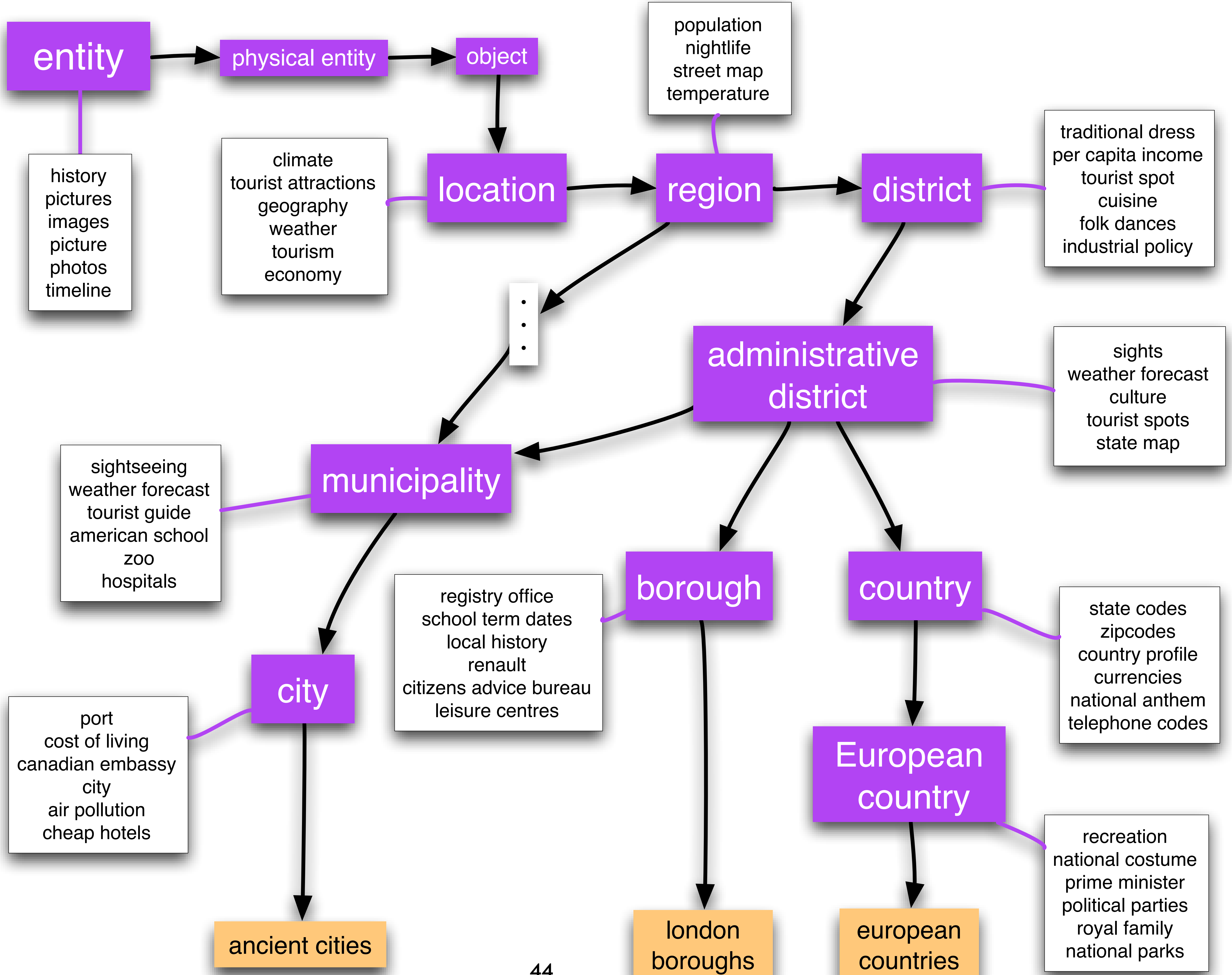



$$DRR = \max \frac{1}{rank(c) \times (1 + PathToGold)}$$

- DRR measures how far each attribute is from its optimal WN node.
- e.g., should "scientific name" be attached to "organism" or "living thing"
- Gold set is constructed by asking raters to give attributes for WN nodes


entity
physical entity
object
whole
living thing
photos
taxonomy
scientific name
reproduction
life cycle
habitat
history
pictures
images
picture
photos
timeline
causal agent
organism
parasite
pathogen
phobia
mortality rate
symptoms
treatment
parasites
bio
autobiography
childhood
bibliography
accomplishments
timeline
person
influences
artwork
style
work
art
technique
entertainer
creator
hairstyle
hairstyles
music videos
songs
new pictures
sexy pictures
performer
artist
self portrait
paintings
famous works
self portraits
painting techniques
famous paintings
filmography
new movies
schedule
new pictures
new pics
actor
influenced
impressionist
the life
's paintings
style of
watercolor
painter
new wallpapers
upcoming movies
baby pictures
latest wallpapers
bollywood
actors
renaissance
painters


entity
physical entity
matter
chemistry
ingredients
chemical structure
dangers
chemical formula
msds
causal agent
substance
history
pictures
images
picture
photos
timeline
pharmacokinetics
mechanism of action
long term effects
pharmacology
contraindications
mode of action
agent
food
recepies
gift baskets
receipes
rdi
daily allowance
fondue recipes
substance
density
uses
physical properties
melting point
chemical properties
chemical structure
fluid
danger
half life
ingredients
side effects
withdrawal symptoms
sexual side effects
drug
liquid
shelf life
nutritional facts
nutrition information
flavors
nutrition
nutritional information
advertisements
sugar content
adverts
brand
nutrition information
storage temperature
drug of abuse
beverage
sugar content
alcohol content
caffeine content
serving temperature
alcohol percentage
shelf life
carbohydrates
carbs
calories
alcohol content
pronunciation
glass
alcohol
grape
vintage chart
grapes
city
food pairings
cheese
drink mixes
apparitions
pitchers
existence
fantasy art
liquor
wine
liquors
red wines

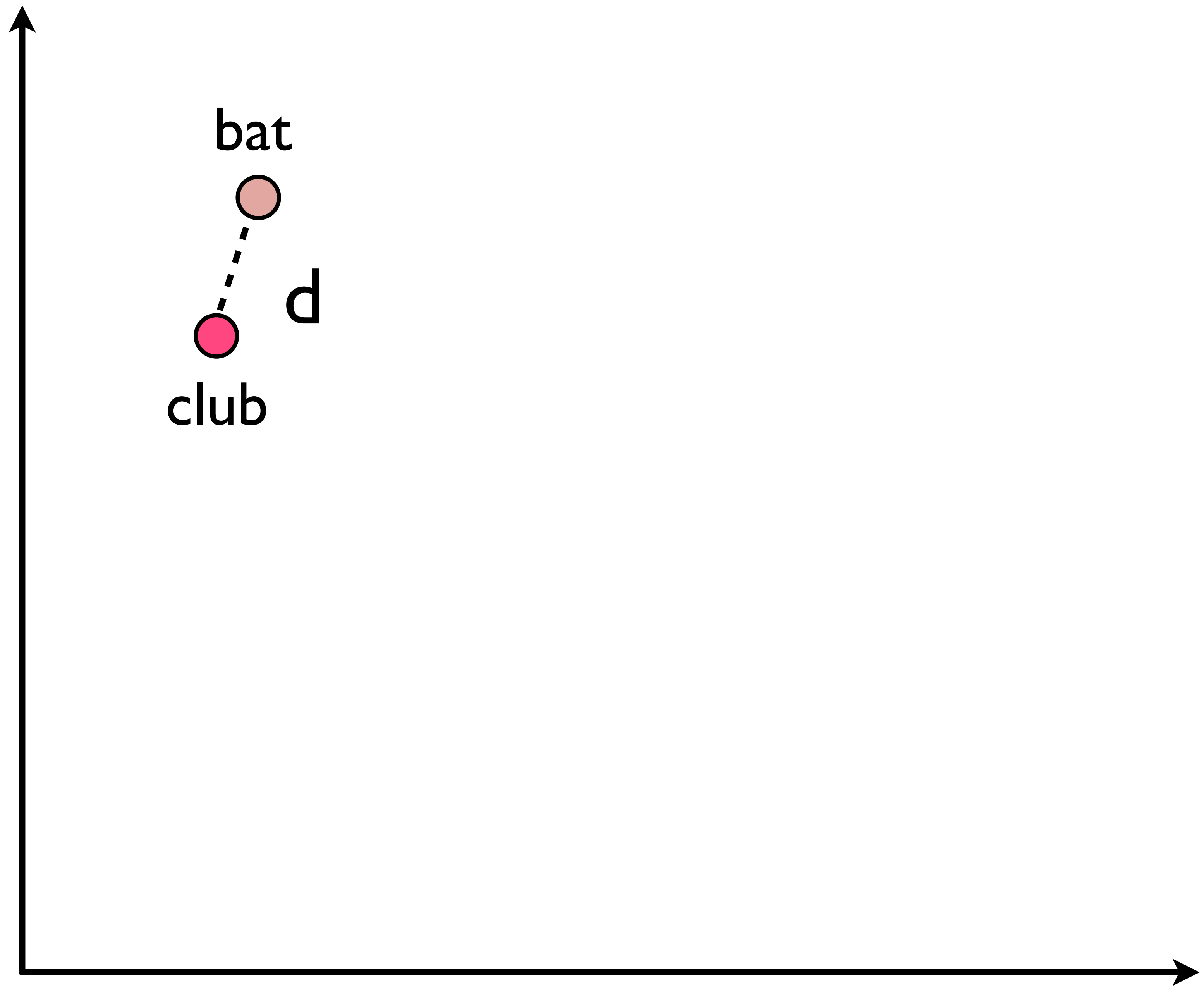

entity
physical entity
object
population
nightlife
street map
temperature
history
pictures
images
picture
photos
timeline
climate
tourist attractions
geography
weather
tourism
economy
location
region
district
traditional dress
per capita income
tourist spot
cuisine
folk dances
industrial policy
administrative district
sights
weather forecast
culture
tourist spots
state map
sightseeing
weather forecast
tourist guide
american school
zoo
hospitals
municipality
borough
country
registry office
school term dates
local history
renault
citizens advice bureau
leisure centres
state codes
zipcodes
country profile
currencies
national anthem
telephone codes
city
port
cost of living
canadian embassy
city
air pollution
cheap hotels
European country
recreation
national costume
prime minister
political parties
royal family
national parks
ancient cities
london boroughs
european countries

# Multi-Prototype Models

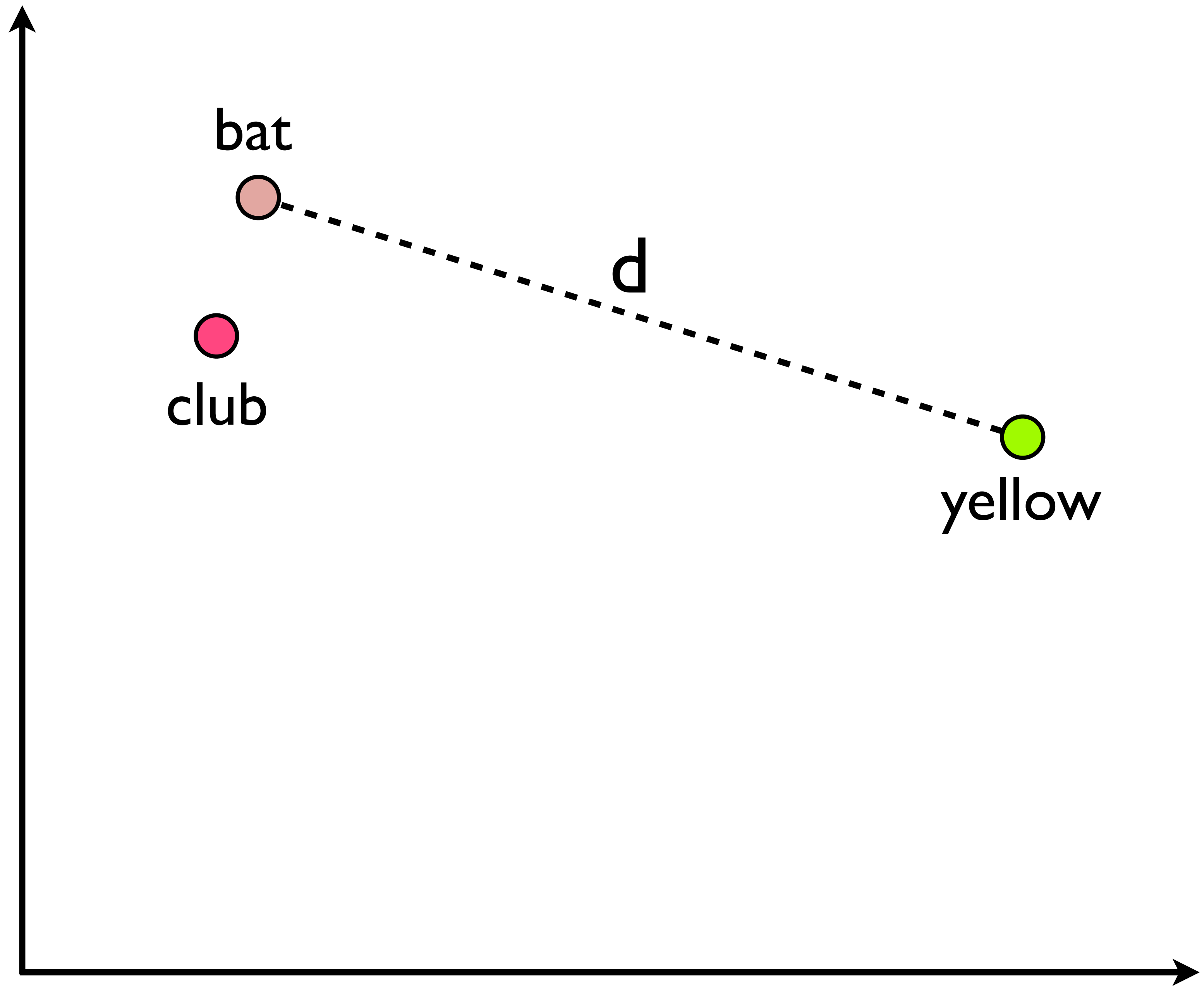
# Vector Space Lexical Semantics

- Represent "meaning" as a point in some high-dimensional space
- Word relatedness correlates with some distance metric
- Attributional: Almuhareb and Poesio (2004), Bullinaria and Levy (2007), Erk (2007), Griffiths et al. (2007), Landauer and Dumais (1997), Padó and Lapata (2007), Sahlgren (2006), Schütze (1997)
- Relational: Moldovan (2006), Pantel and Pennacchiotti (2006), Turney (2006)

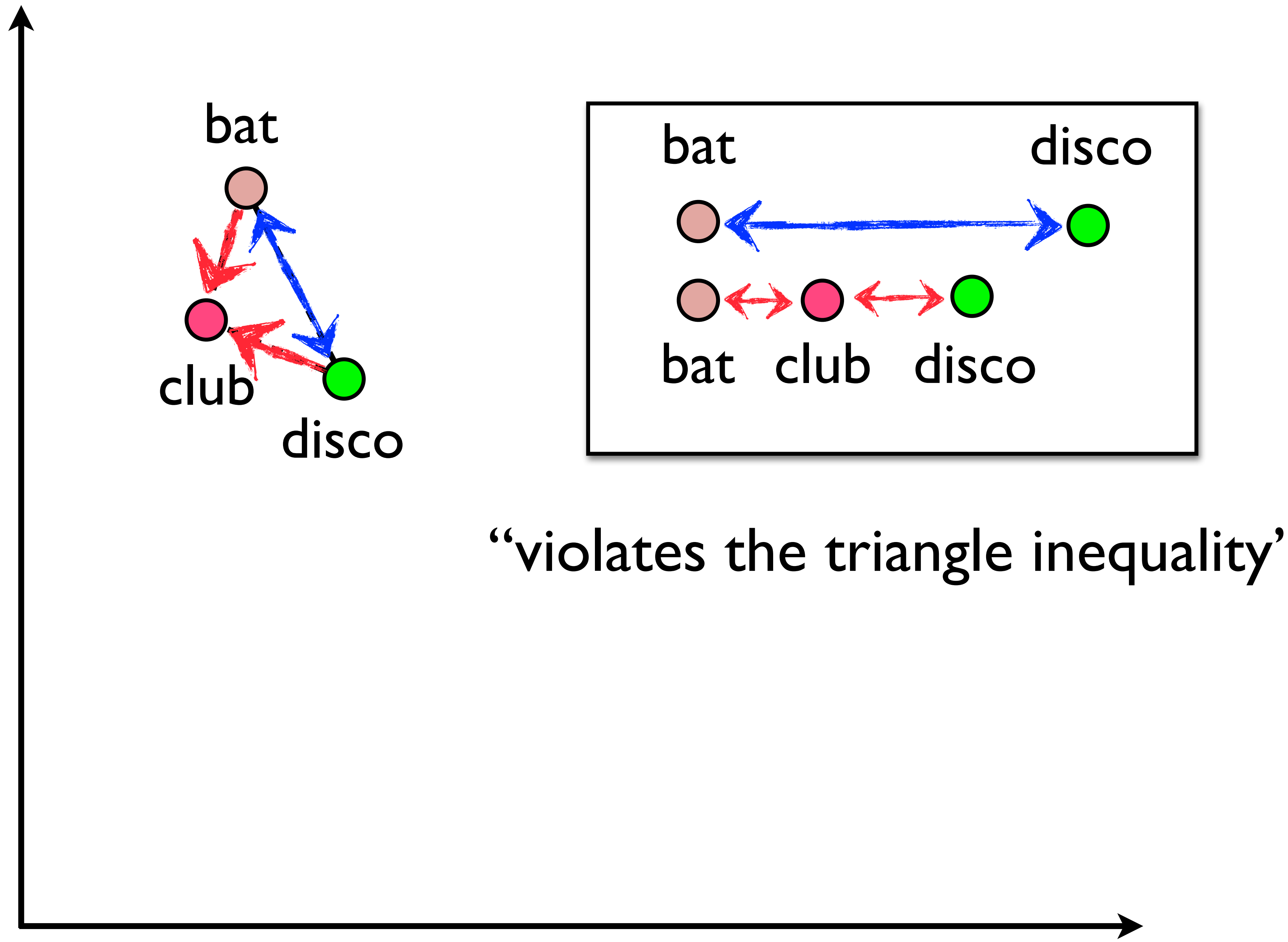
$\Omega =$



$$\Omega =$$



$\Omega =$

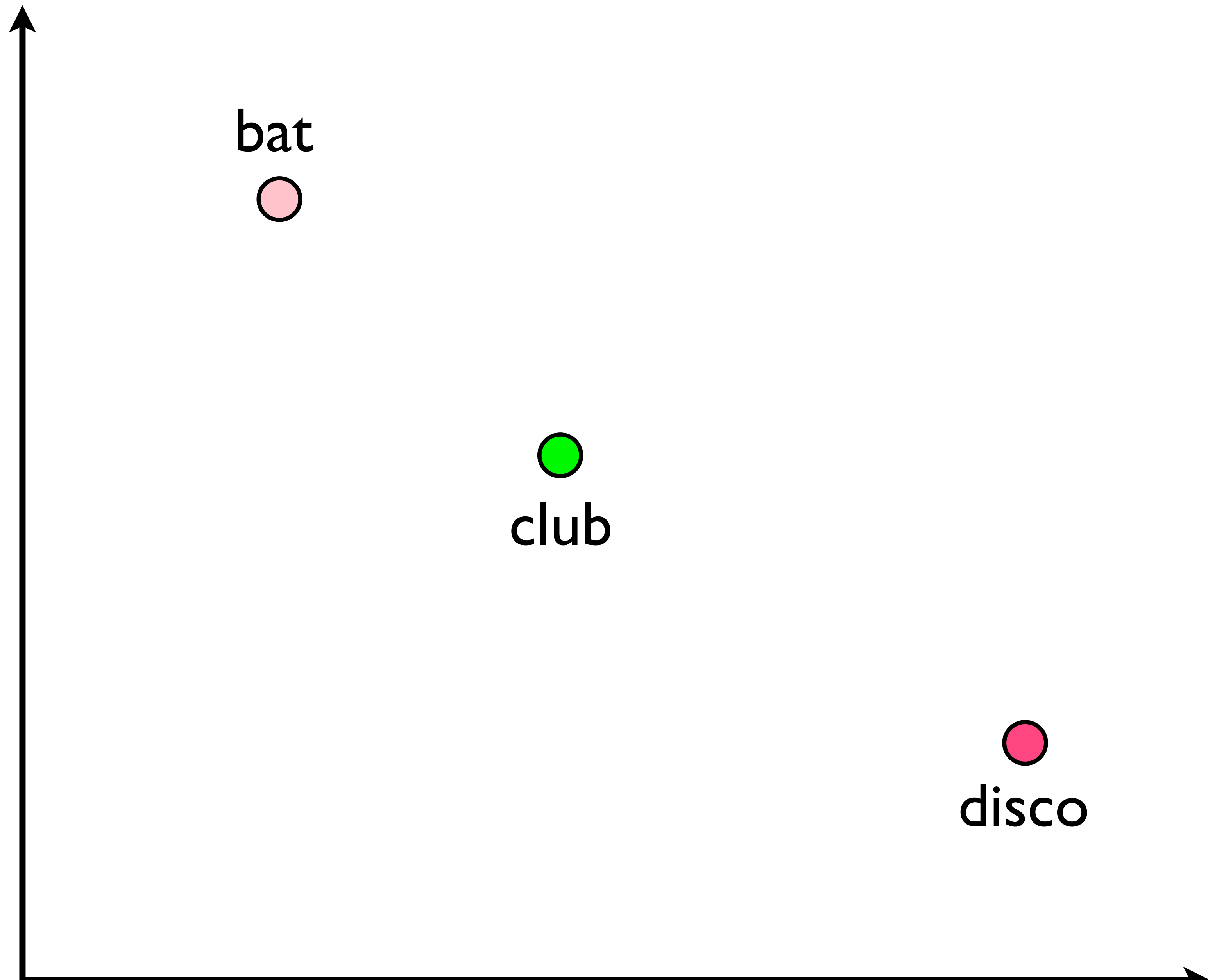



- Any inner product space; e.g. “dense” semantic spaces like LSA

Tversky and Gati (1982), Griffiths et al. (2007)

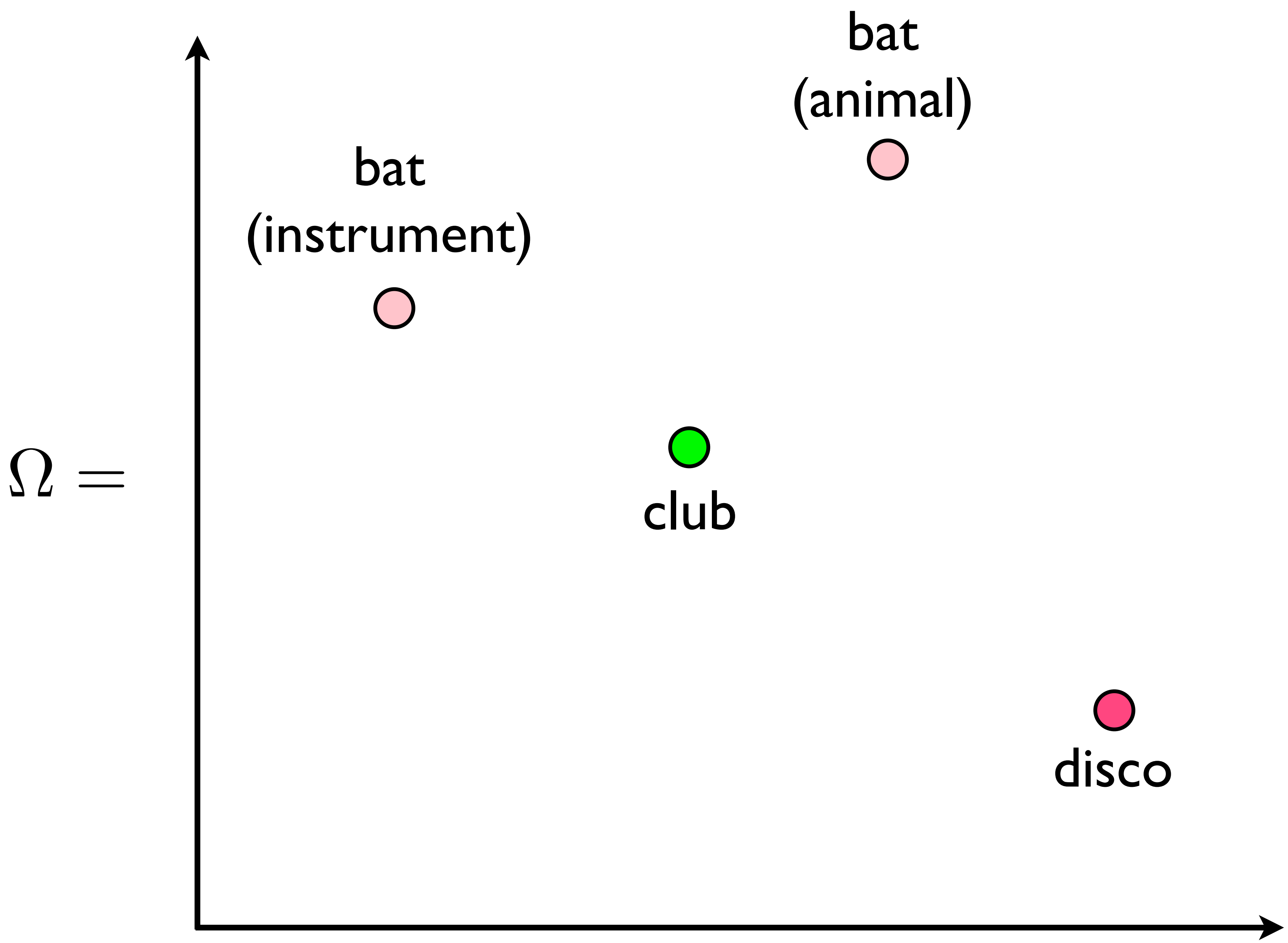
# Using multiple prototypes

$\Omega =$




- Similar to unsupervised Word Sense Discovery, e.g. Pantel and Lin (2002), Schütze (1998), Yarowsky (1995)

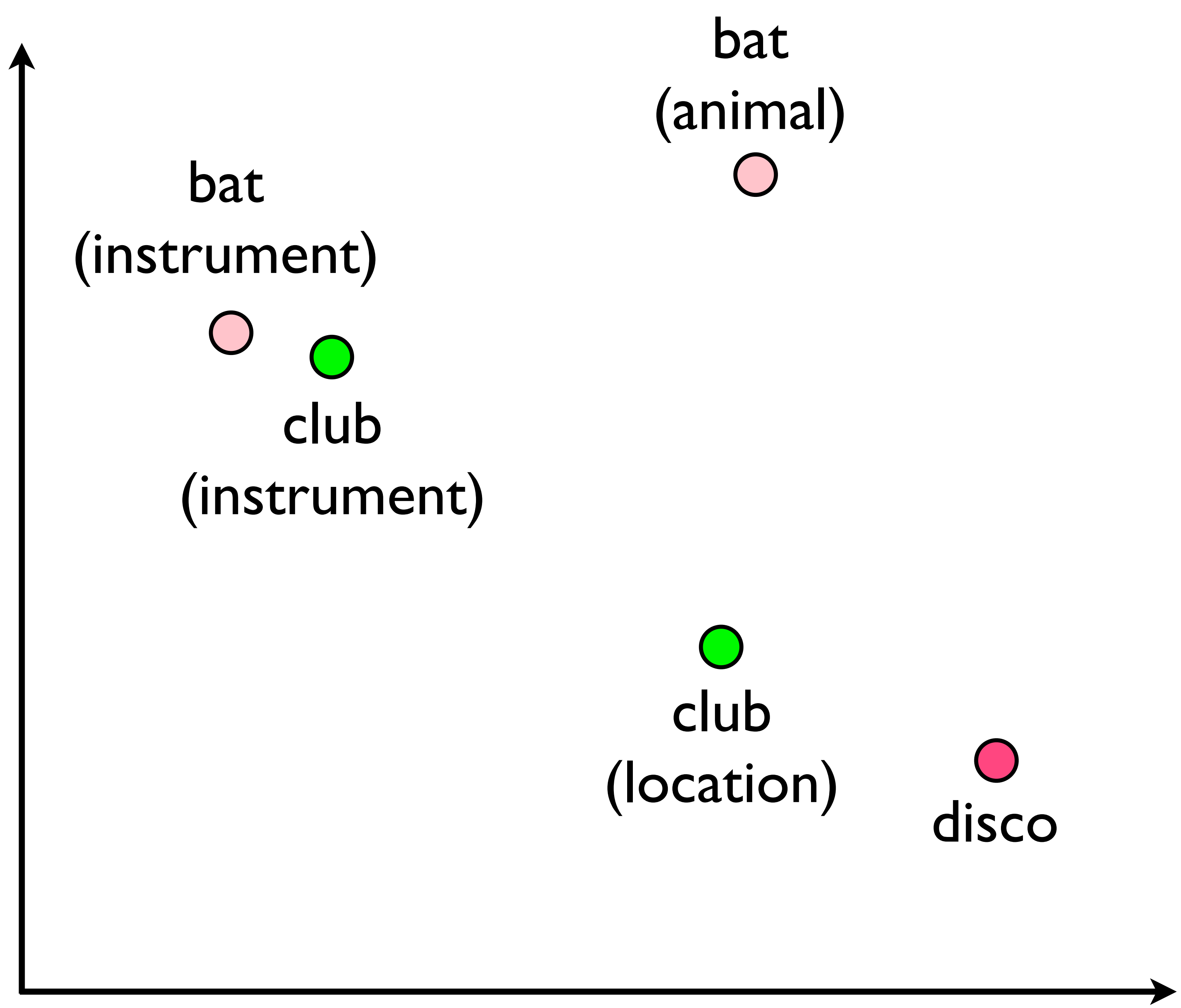
# Using multiple prototypes




- Similar to unsupervised Word Sense Discovery, e.g. Pantel and Lin (2002), Schütze (1998), Yarowsky (1995)

# Using multiple prototypes

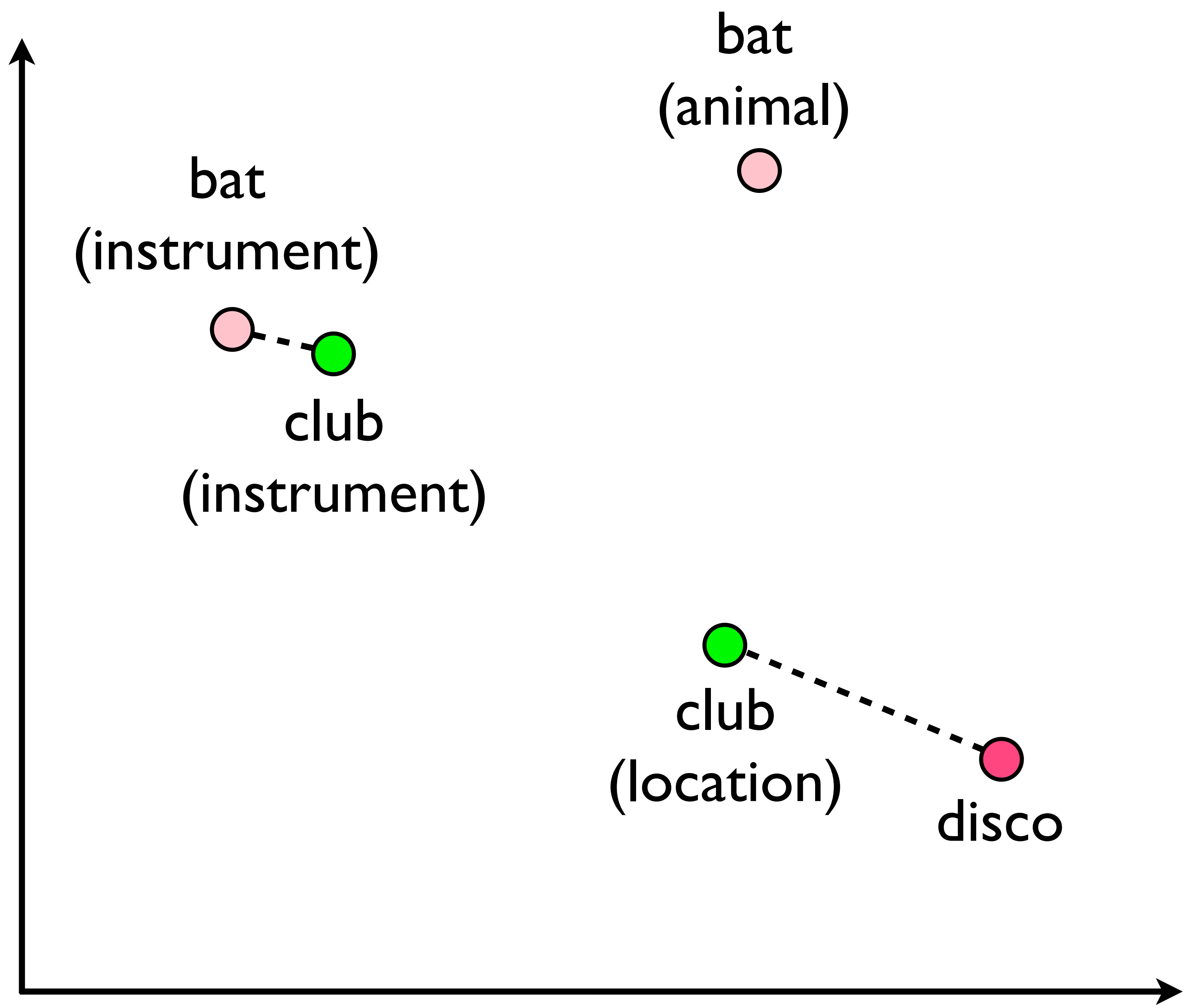
$\Omega =$




- Similar to unsupervised Word Sense Discovery, e.g. Pantel and Lin (2002), Schütze (1998), Yarowsky (1995)

# Using multiple prototypes

$$\Omega =$$

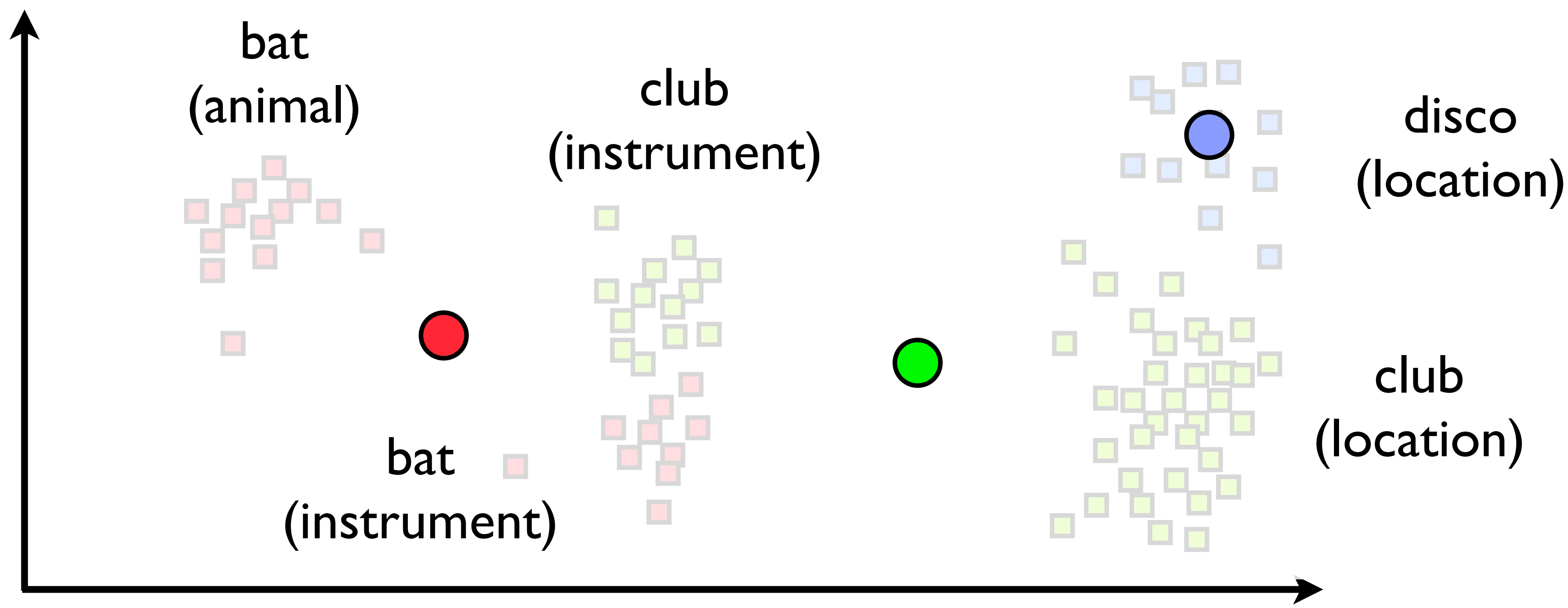



- Similar to unsupervised Word Sense Discovery, e.g. Pantel and Lin (2002), Schütze (1998), Yarowsky (1995)

# Some practical benefits

- "Meaning" is a mixture over prototypes, capturing polysemy and thematic variation.
- Can exploit contextual information to refine word similarity computations:
  - e.g., is "the bat flew out of the cave" similar to "the girls left the club" ?
- "Senses" are thematic and very fine-grained
  - e.g., the *hurricane* sense of *position*

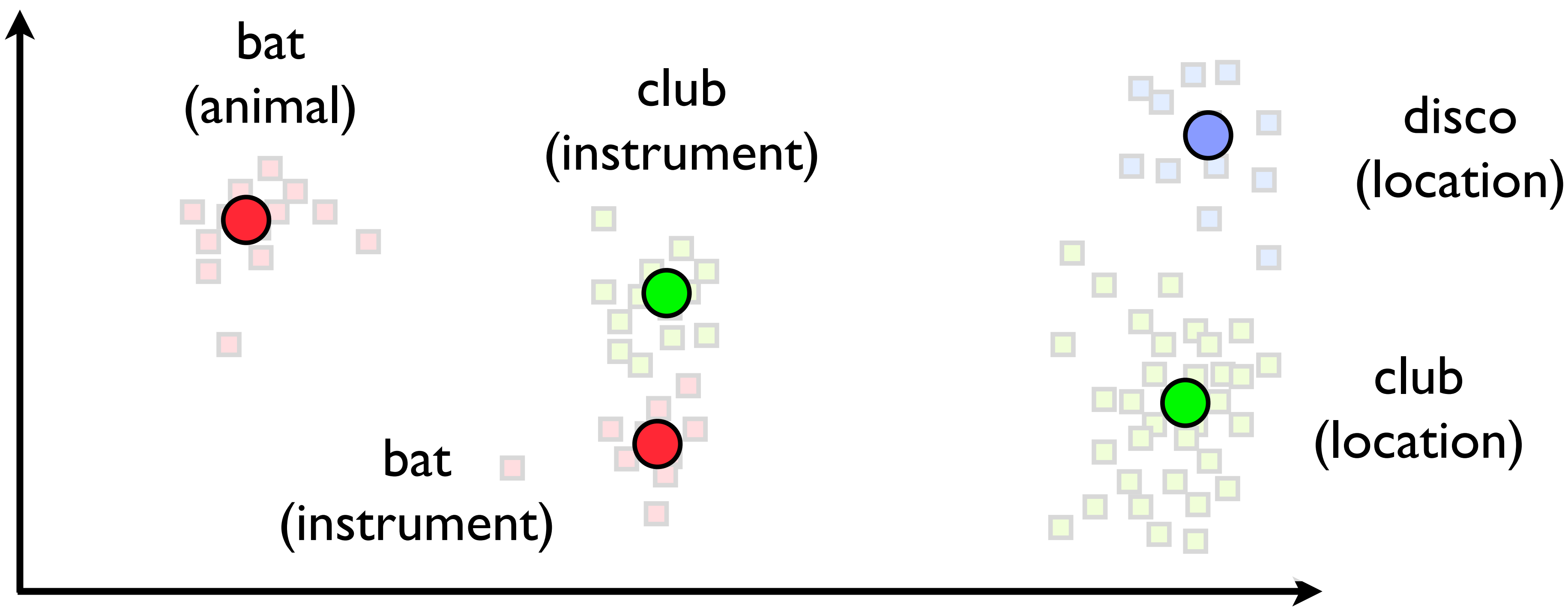
# Single Prototype ↔ Multi-Prototype ↔ Exemplar

$\Omega =$ 


- Find the centroid of the individual word occurrences
- Conflates senses

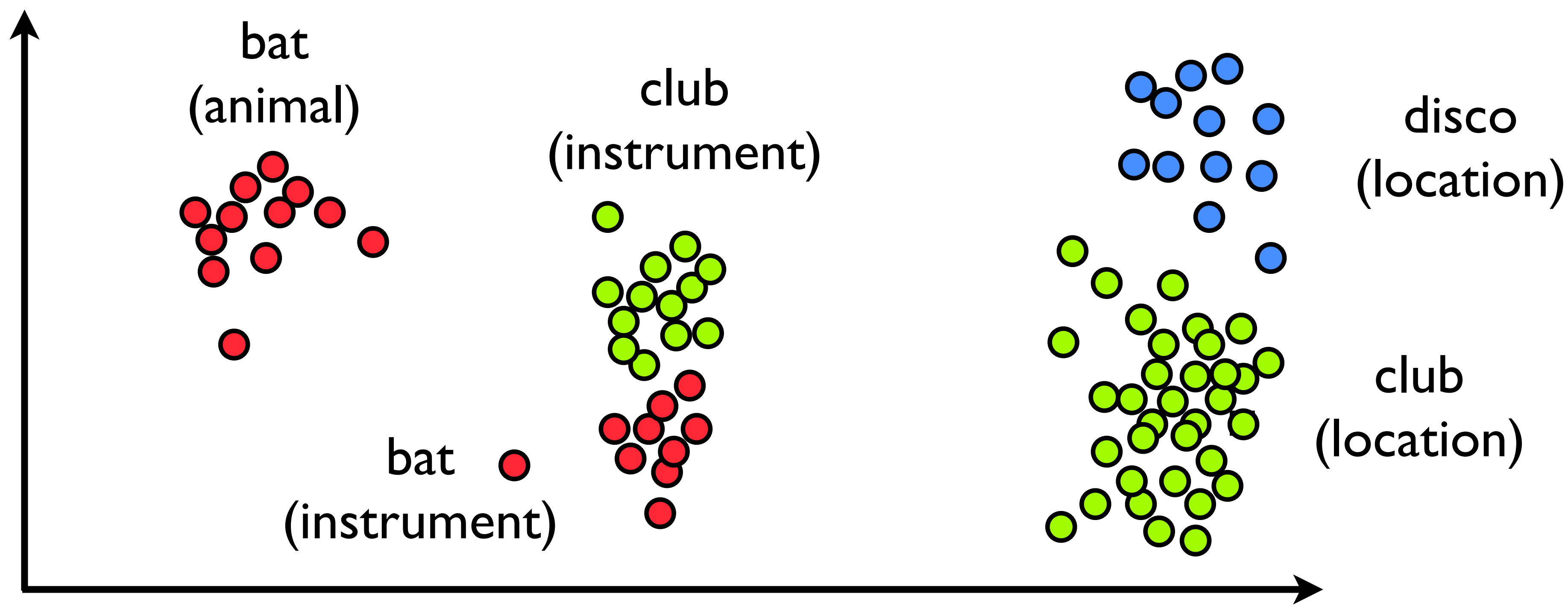
# Single Prototype ↔ Multi-Prototype ↔ Exemplar

$\Omega =$




- Essentially just clustering word occurrences
- Doesn't find lexicographic senses; captures contextual variance directly.

# Single Prototype ↔ Multi-Prototype ↔ Exemplar

$\Omega =$ 


- Just treat all occurrences as an ensemble representing meaning.
- Compute similarity as the average of the K most similar pairs.
- Heavily influenced by noise, but captures more structure

Erk (2007), Vandekerckhove et al. (2009)

# Feature Engineering / Weighting

- Choosing an embedding vector space:
  - features (unigram, bigram, collocation, dependency, ...)
  - feature weighting (t-test, tf-idf, $\chi^2$, MI, ...)
  - metric / inner product (cosine, Jaccard, KL, ...)
- The multi-prototype method is essentially agnostic to these implementation details

Curran (2004)

# Feature Engineering / Weighting

- Choosing an embedding vector space:
  - features (unigram, bigram, collocation, dependency, ...)
  - feature weighting (t-test, tf-idf, $\chi^2$, MI, ...)
  - metric / inner product (cosine, Jaccard, KL, ...)
- The multi-prototype method is essentially agnostic to these implementation details

Curran (2004)

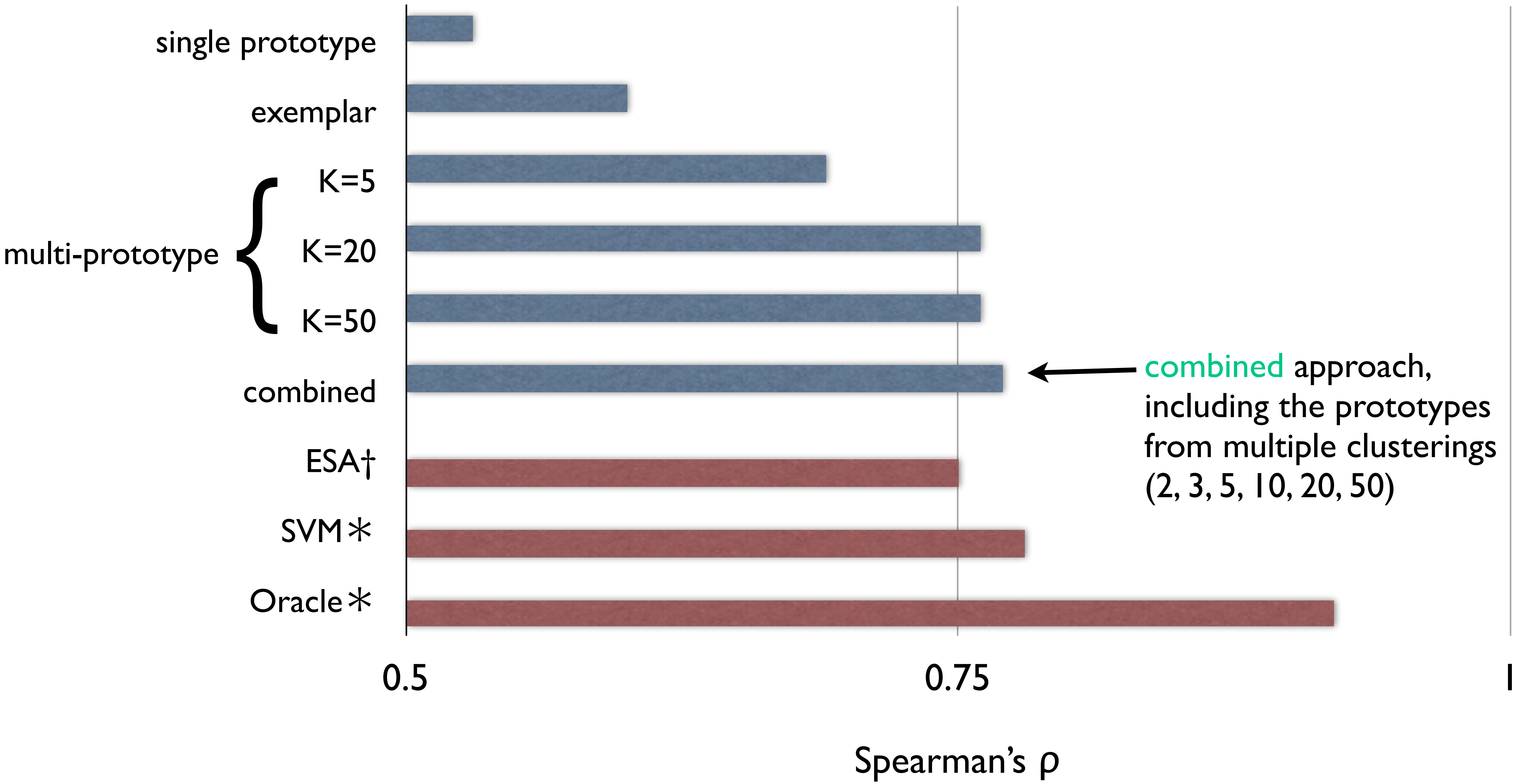
# Experimental setup

- Wikipedia as the base textual corpus (2.8M articles, 2B words)
- Evaluation:

1. WordSim-353 collection (353 word pairs with ~15 human similarity judgements each) (Finkelstein et al. (2002)); using Spearman's rank correlation ( Agirre et al. (2009))

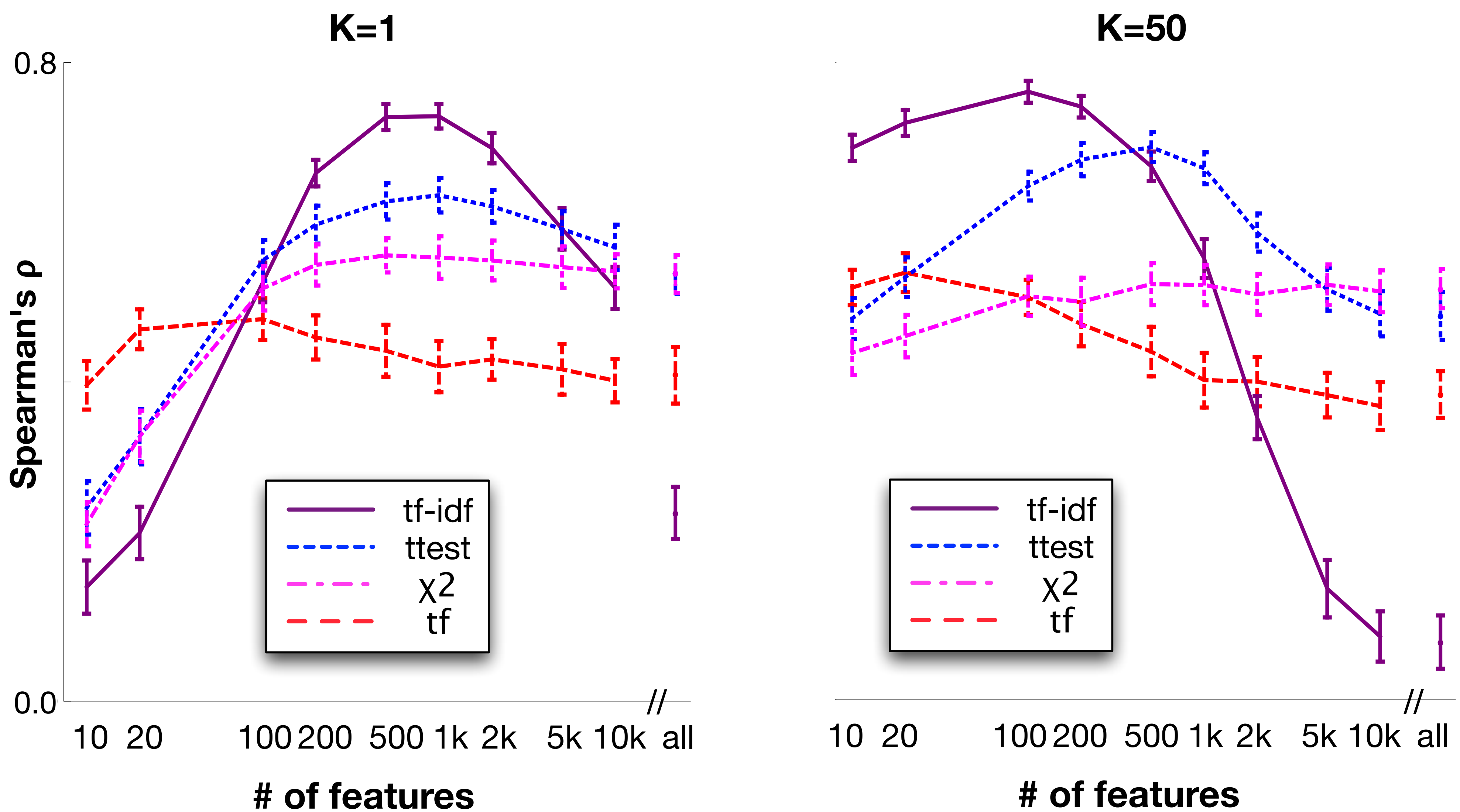
2. Predicting related words; human raters from Amazon Mechanical Turk

# Results: WordSim-353 Correlation




†Gabrilovich and Markovitch (2007), ＊ Agirre et al. (2009)

# Feature pruning is really important




- Computed correlation over a number of different feature weightings with different amounts of pruning

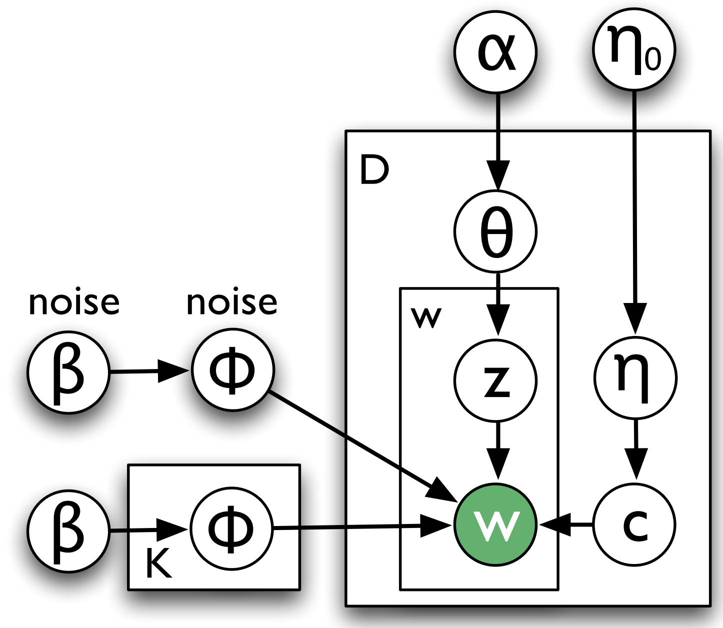
- Pruning = cut out all but the top X features by weight from vector

# Proposed Work

- Feature weighting and pruning is important
- The topic model work is basically feature re-weighting
- Can we use topic models more generally to perform feature selection?

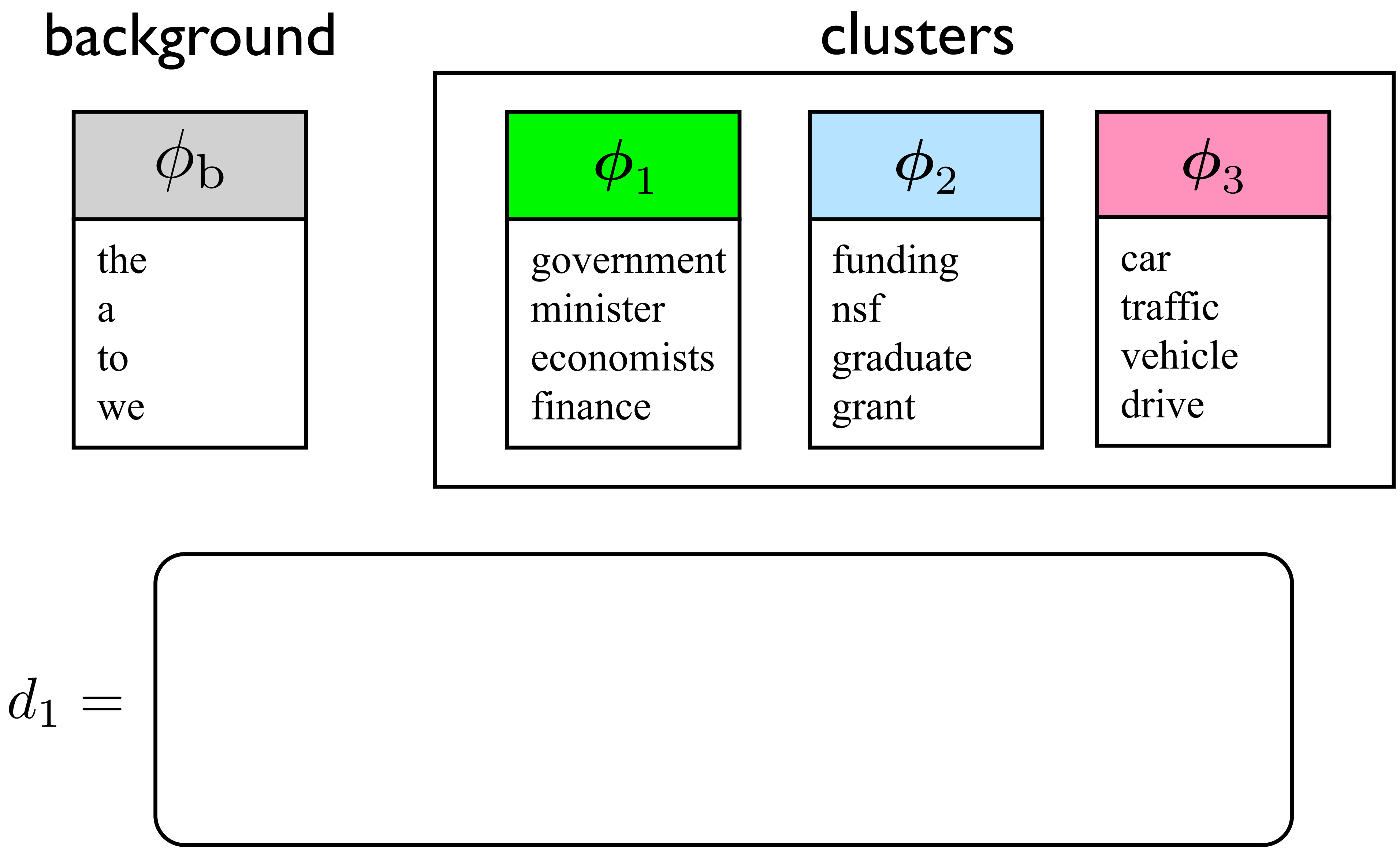
# Explicit feature selection via *tiered clustering*

$$
\begin{array}{rcllr}
\eta_d|\eta_0 & \sim & \text{Beta}(\eta_0) & d \in D, & \text{(noise prop)} \\
\boldsymbol{\phi}_k|\boldsymbol{\beta} & \sim & \text{Dirichlet}(\boldsymbol{\beta}) & k \in K, & \text{(clusters)} \\
\boldsymbol{\phi}_{\text{noise}}|\boldsymbol{\beta}_{\text{noise}} & \sim & \text{Dirichlet}(\boldsymbol{\beta}_{\text{noise}}) & & \text{(noise)} \\
\boldsymbol{\theta}_d|\boldsymbol{\alpha} & \sim & \text{Dirichlet}(\boldsymbol{\alpha}) & d \in D, & \text{(cluster prop)} \\
c_d|\boldsymbol{\theta}_d & \sim & \text{Mult}(\boldsymbol{\theta}_d) & d \in D, & \text{(cluster ind)} \\
z_{i,d}|\eta_d & \sim & \text{Bernoulli}(\eta_d) & i \in |\mathbf{w}_d|, & \text{(noise ind)} \\
w_{i,d}|\boldsymbol{\phi}_{c_d}, z_{i,d} & \sim & \begin{cases} \text{Mult}(\boldsymbol{\phi}_{\text{noise}}) \\ \quad (z_{i,d} = 1) \\ \text{Mult}(\boldsymbol{\phi}_{c_d}) \\ \quad \text{(otherwise)} \end{cases} & i \in |\mathbf{w}_d|, & \text{(words)}
\end{array}
$$



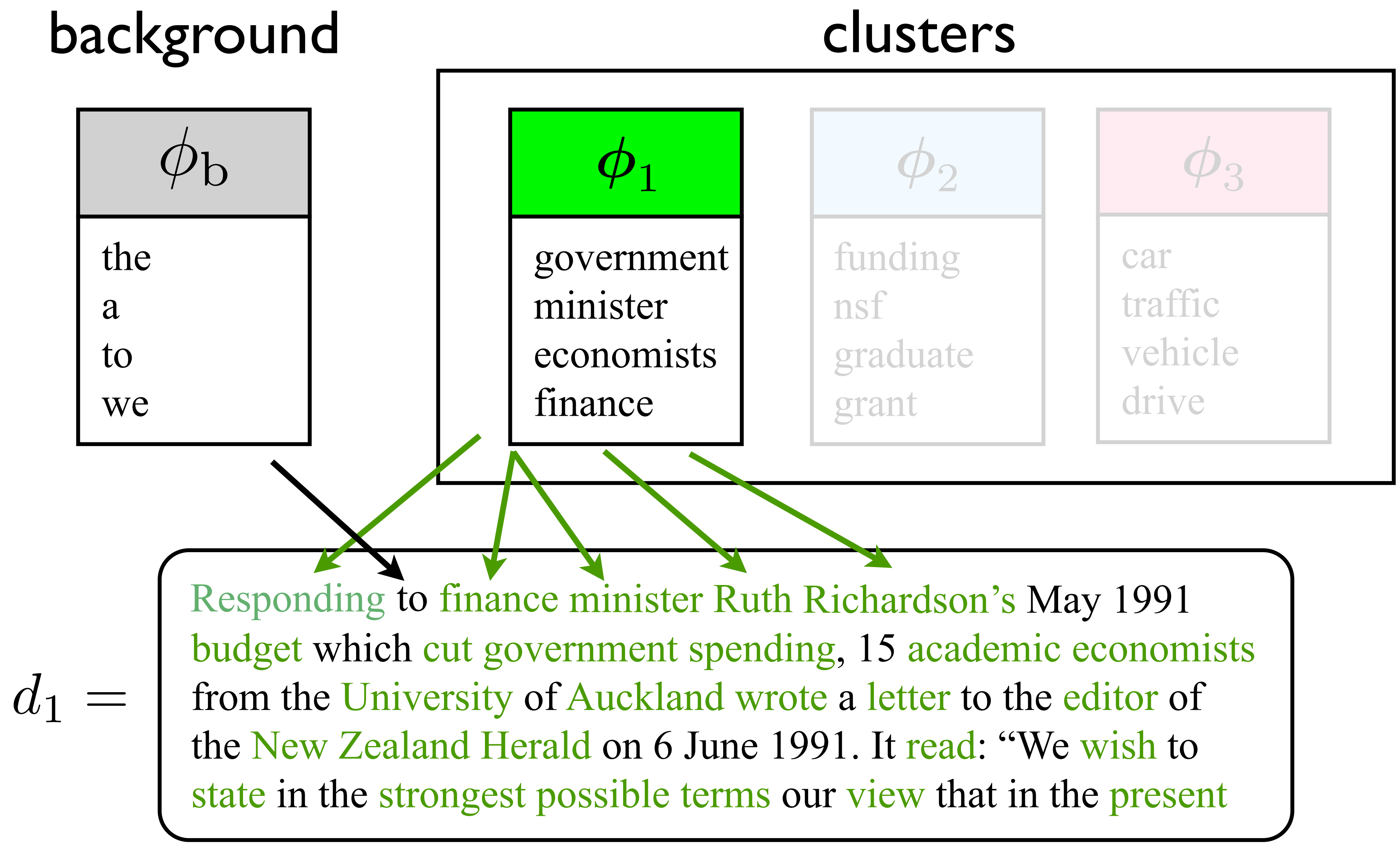

- Generalization of feature-selective clustering
- Uses topic modeling for "soft" feature selection

# Explicit feature selection via *tiered clustering*




- Generalization of feature-selective clustering
- Uses topic modeling for "soft" feature selection

# Explicit feature selection via *tiered clustering*




- Generalization of feature-selective clustering
- Uses topic modeling for "soft" feature selection

# From feature selection to multiple clustering

- Really we don't want pruning, we want multiple clusterings; modeling conditional feature noise
- Cross-cutting categorization in psychology
- Remember: people use multiple categorization systems
  - (e.g. situational vs. taxonomic categorization of food)

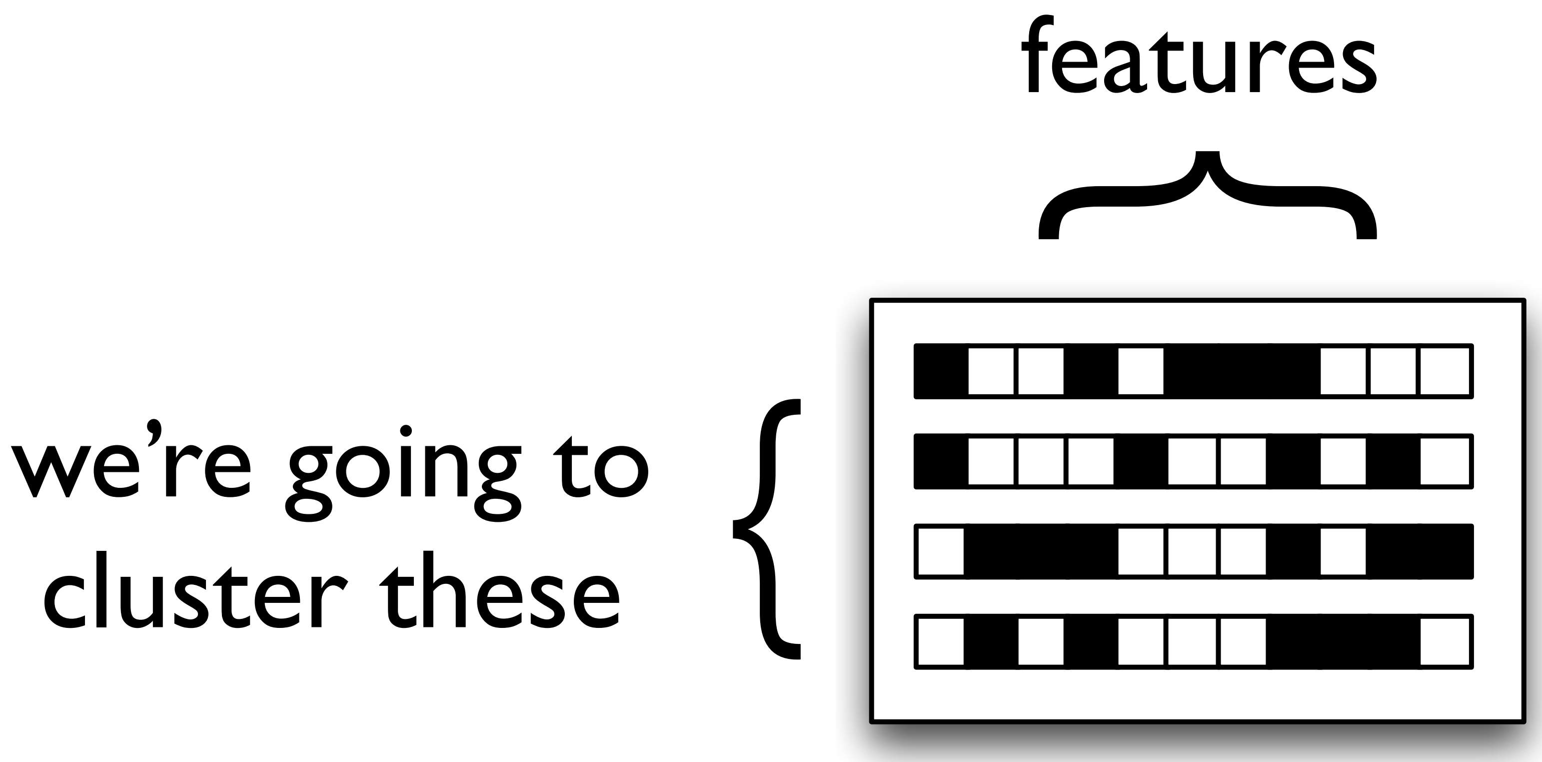
# Another hypothesis

- Cross-cat structure finds coherent subsets of features

- These feature subsets capture specific contextual generalization

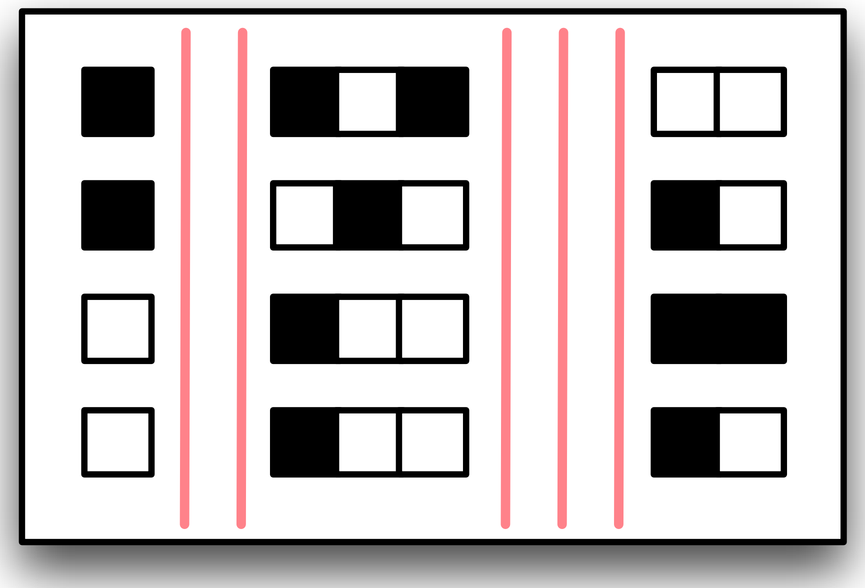
- Each clustering implicitly defines a different relation latent in the data

# Building a cross-cutting categorization model




- This is the clustering step in the multi-prototype model
- or, e.g. clustering concepts

# Building a cross-cutting categorization model



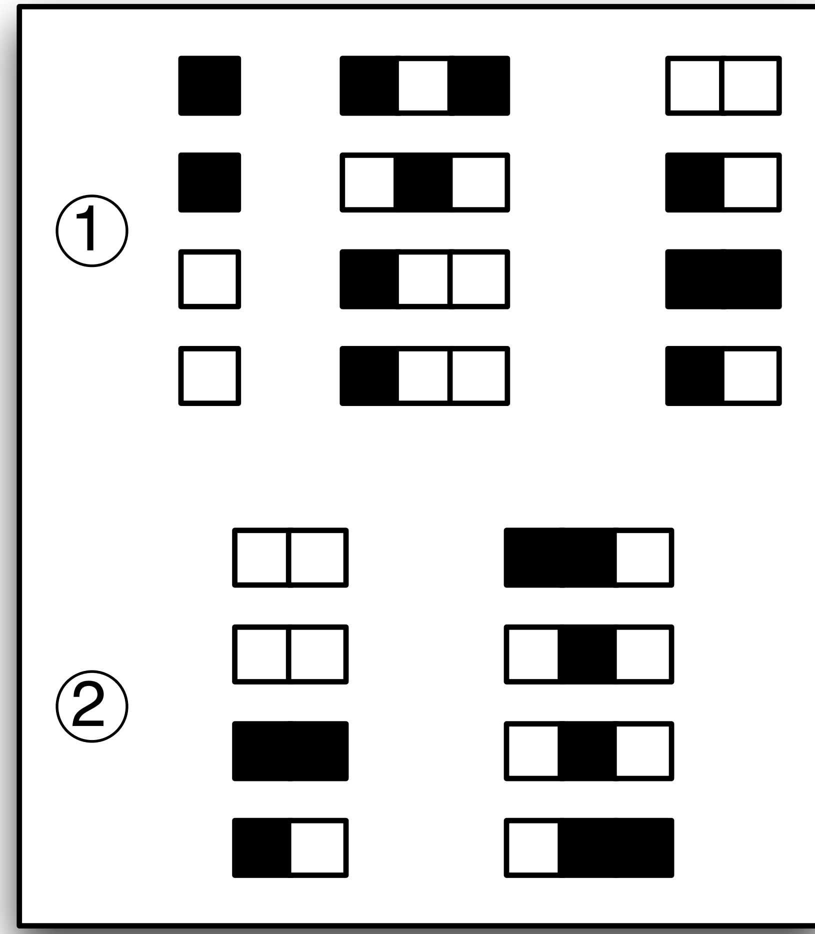
feature selective clustering

- Cluster using only a subset of the available features
- More robust to feature correlation / noise

# Cross-cutting categorization

E.g., (1) contains syntactic features and (2) contains topical features

or, e.g. (1) is occupations and (2) is locations




cross-categorization / multiple views

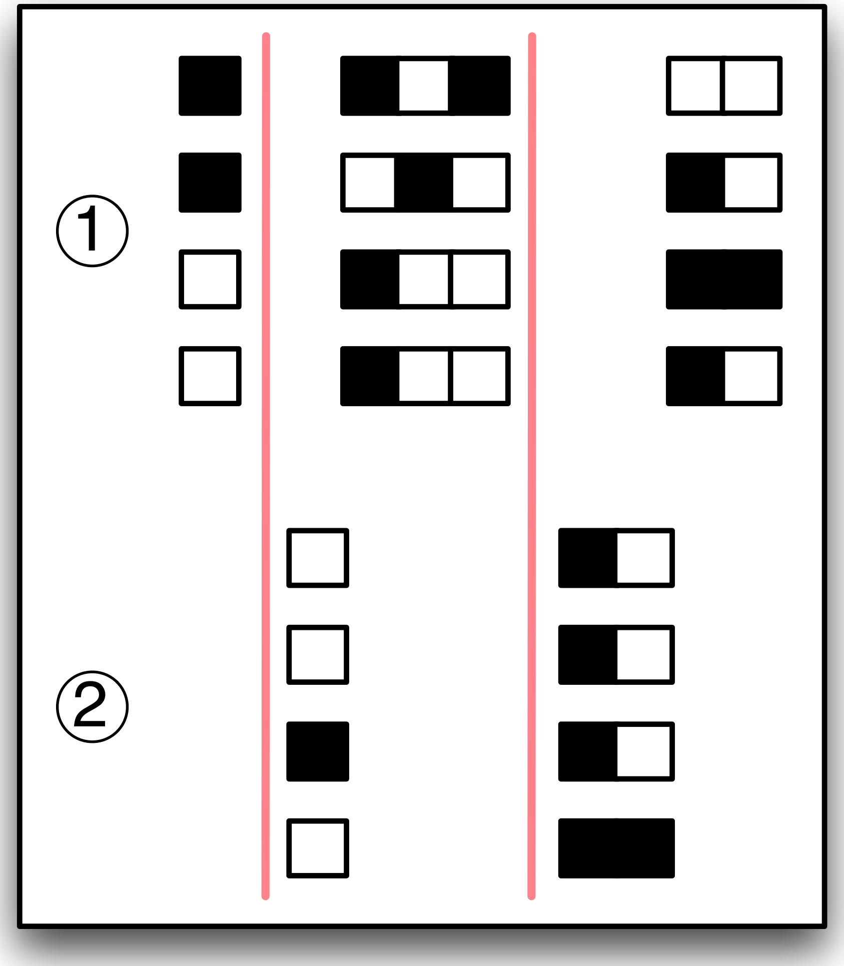
- Divide features among mutually-exclusive views; features in the same view highly covary

- Each view captures a different way of clustering the data

# Cross-cutting categorization

E.g., (1) contains syntactic features and (2) contains topical features

or, e.g. (1) is occupations and (2) is locations




feature selection + cross-categorization

- Divide data among views;
- Remove features that fail to yield any consistent clustering of the data

# Cross-cat on the Dartmouth Health Atlas

| "QoC Scores" 5 categories | "Long Term Care" 7 categories | "Hospice Care" 11 categories | "Home Care" 5 categories | "Specialist vs. PC" 10 categories | "Skilled Nursing" 18 categories | "Equipment" 7 categories |
|---|---|---|---|---|---|---|
| Composite Quality Score | Medicare $ / Decedent on LTC | Medicare $ / Decedent on Hospice Care | Medicare $ / Decedent on Home Care | Ratio of Specialist to PC FTEs | SNF Beds / 1000 Decedents | Medicare $ / Decedent on Durable Equipment |
| AMI Score | Medicare $ / Decedent on Ambulance | Percent of Deaths Occurring in Hospice | Home Health Agency Visits / Decedent | Ratio of Specialist to PC Visits/ Decedent | SNF Days / Decedent | Durable Equipment Copay / Decedent |
| CHF Score | | Hospice Days / Decedent | | | | |
| Pneumonia Score | | | | | | |

| "Misc. Spending" 9 categories |
|---|
| Medicare $ / Decedent on Outpatient Care |
| Medicare $ / Decedent on Other Care |
| Medicare Part B $ / Decedent on Procs. |
| Medicare Part B $ / Decedent on Imaging |
| Medicare Part B $ / Decedent on Tests |
| Medicare Part B $ / Decedent on Other |
| Total Copay / Decedent |
| Physician Services Copay / Decedent |

- 4273 hospitals; 74 features including quality scores and spending measurements
- Each view captures a set of variables correlated with each other, decorrelated from the other views

Mansinghka et al. (2009)

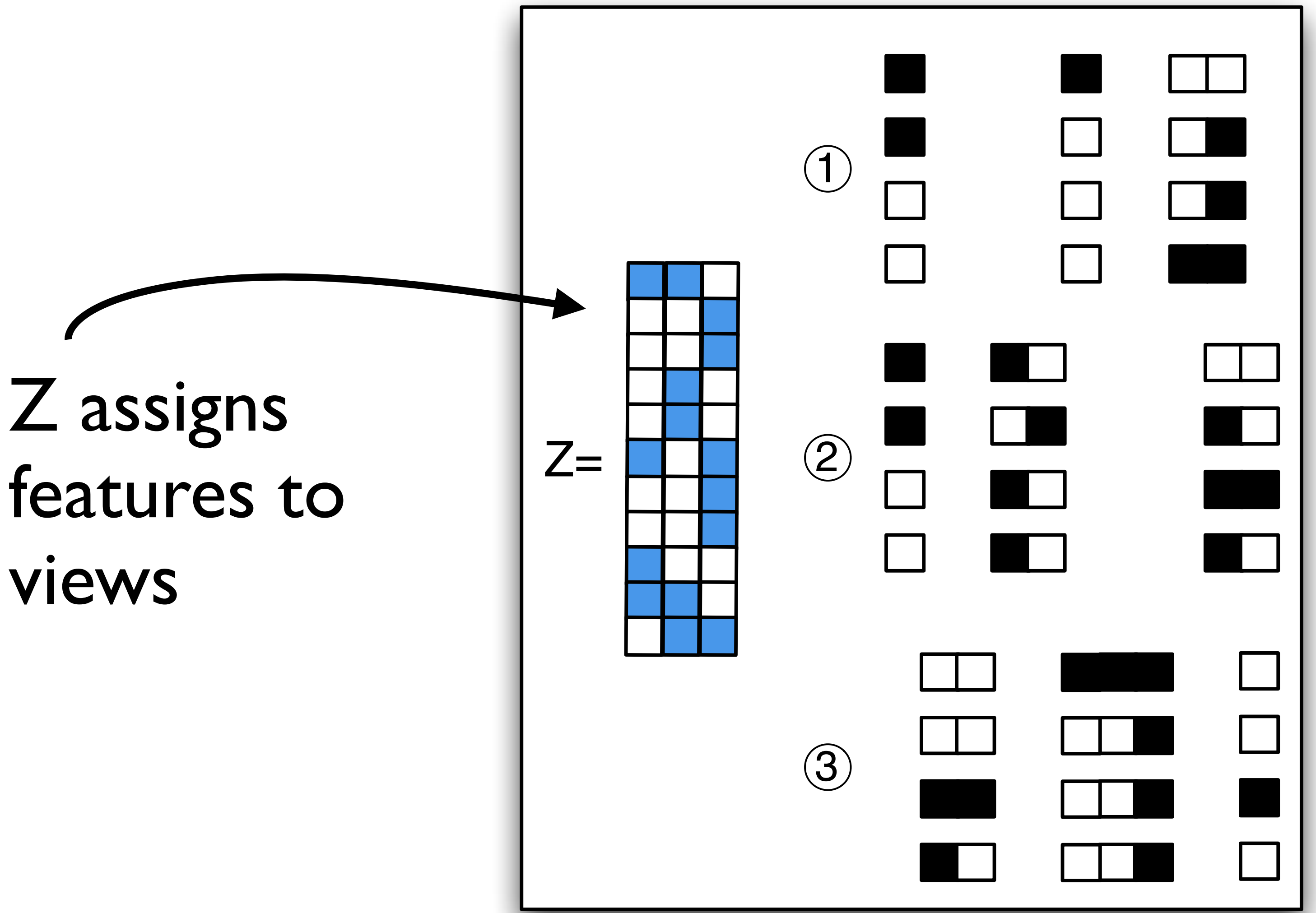
# Cross-cat on the Dartmouth Health Atlas

| "QoC Scores" 5 categories | "Long Term Care" 7 categories | "Equipment" 7 categories | "Misc. Spending" 9 categories |
|---|---|---|---|
| Composite Quality Score | Medicare $ / Decedent on LTC | Medicare $ / Decedent on Durable Equipment | Medicare $ / Decedent on Outpatient Care |
| AMI Score | Medicare $ / Decedent on Ambulance | Durable Equipment Copay / Decedent | Medicare $ / Decedent on Other Care |
| CHF Score | | | Medicare Part B $ / Decedent on Procs. |
| Pneumonia Score | | | Medicare Part B $ / Decedent on Imaging |
| | | | Medicare Part B $ / Decedent on Tests |
| | | | Medicare Part B $ / Decedent on Other |
| | | | Total Copay / Decedent |
| | | | Physician Services Copay / Decedent |

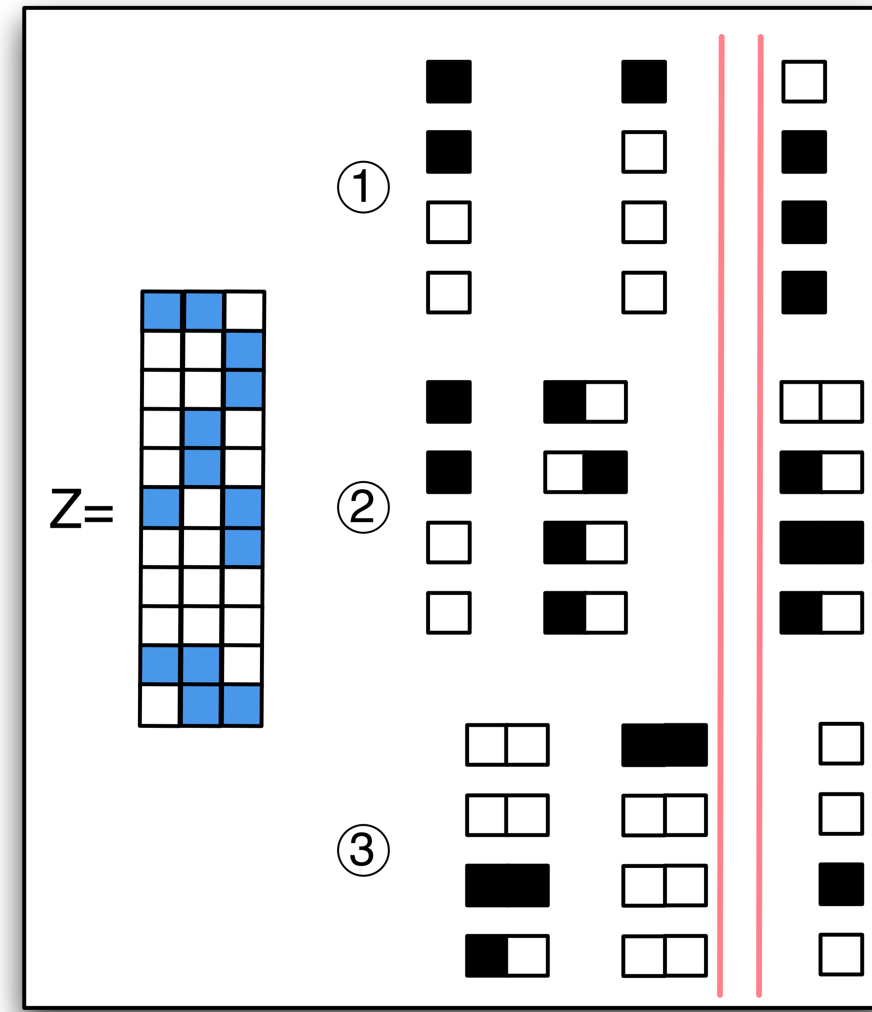
- No connection between quality of care, hospital size and spending.
- Increases in long-term care causes increase ambulance costs (e.g. an at-home mishap)
- Clustering alone misses this cross-cutting structure

Mansinghka et al. (2009)

# Factorial feature allocation



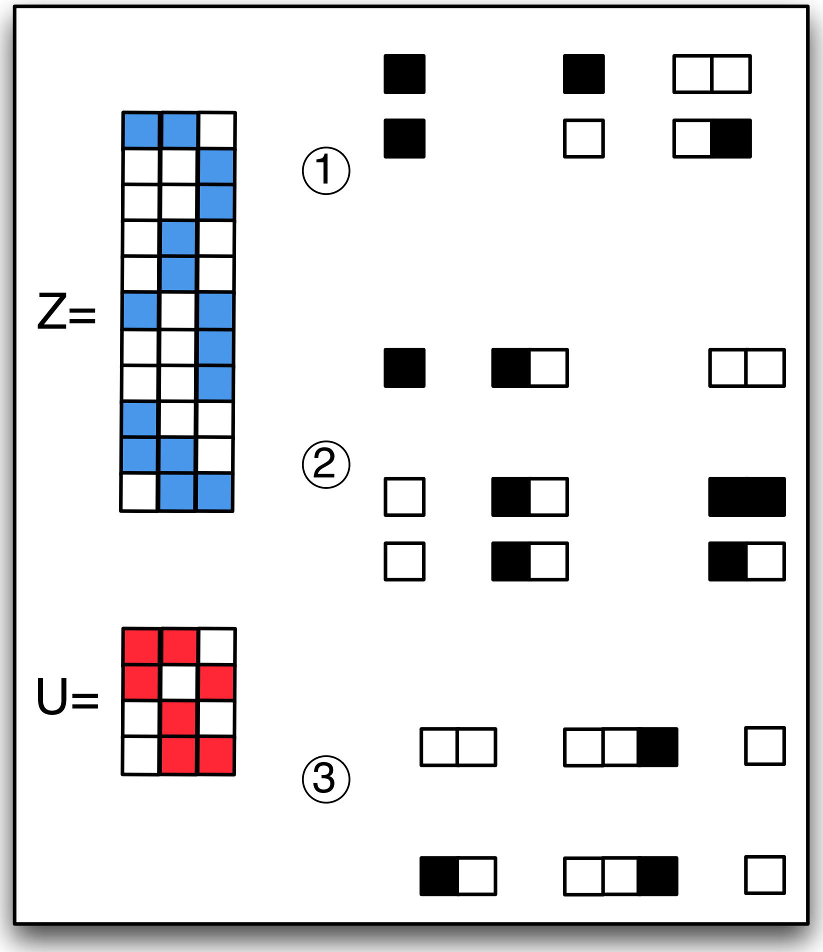

factorial feature allocation




+ feature selection

- Some features might be useful in multiple views
- Cross-cat cannot capture such factorial structure

# Factorial feature and data allocation




factorial feature+data allocation

E.g., organizing animals taxonomically we might want to exclude "fictional ducks", but organizing by physical characteristics we might include them.

- View-dependent outlier detection
- Certain data points might be outliers in certain views, but not in others

# What is this model good for?

# Open, implicit relation extraction via cross-cat

- Each view in Cross-cat characterizes a relation implicitly (i.e. a coherent set of dimensions capturing variance between objects)
- These relations are unlabeled, but go beyond simple "word relatedness" or "word similarity"

Banko and Etzioni (2008)

# Application: Associative anaphora resolution

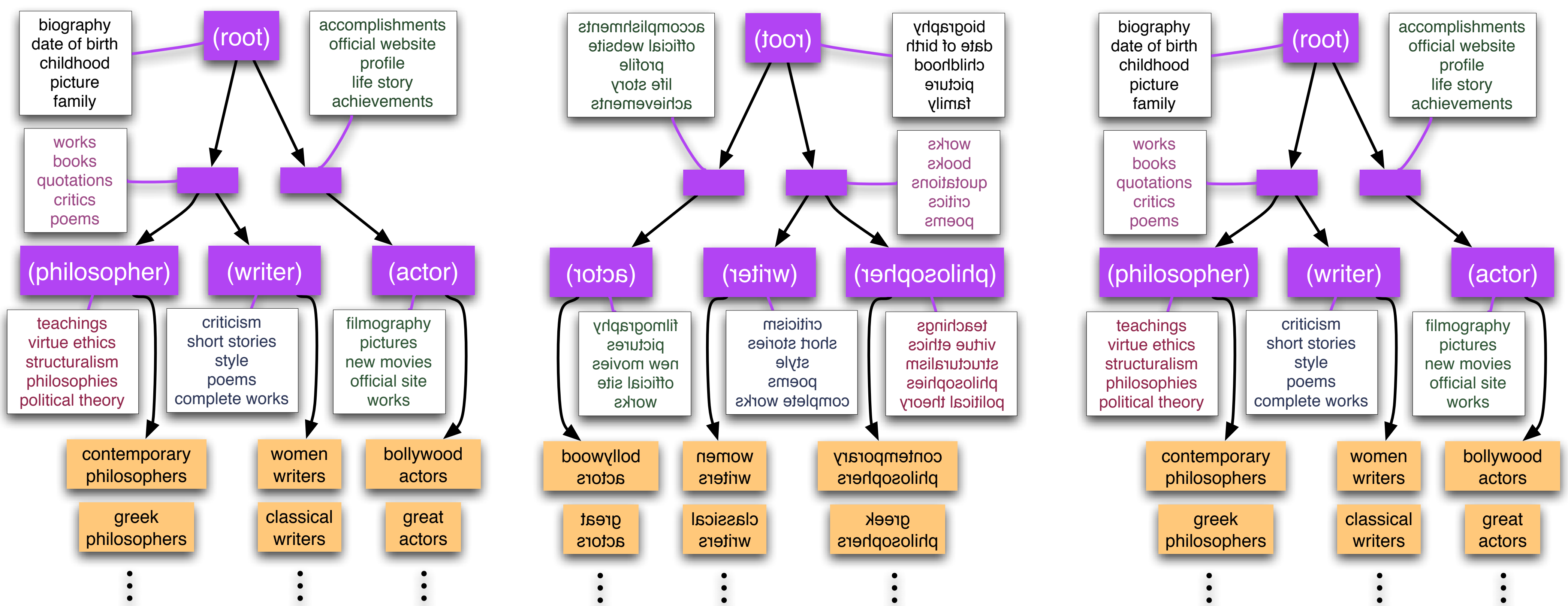
1. Once she saw that all **the tables**$_{(\rightsquigarrow 1)}$ were taken and **the bar**$_{(\rightsquigarrow 1)}$ was crowded, she left **the restaurant**$_{(1)}$.

2. Shares of **AAPL**$_{(2)}$ closed at $241.19. **Volatility**$_{(\leftsquigarrow 2)}$ was below the 10-day moving average.

Charolles (1999), Bunescu (2003), Poesio et al. (2004)

# Application: Hierarchical cross-categorization

View 1 View 2 View 3




- Automate the construction of DAG-structured ontologies (mixture of trees)
- Really want mixtures of local ontologies

# Other applications

- Selectional preference of verbs, adjectives, etc
- Paraphrase acquisition
- Cross-lingual attribute extraction / property generation (are concept categorization systems conserved cross-culturally?)
- Twitter; accounting for rich topical tag structure

# Summary

- Distributional lexical semantics models cannot adequately capture the richness of human concept organization.
- Cross-cutting categorization is a coherent, tractable framework for addressing this issue
- Can broaden the scope of applications for lexical semantics
- ◉ Didn't touch on scalability, but yes, it is

# Thanks!

# Questions?