# Lexical Semantics II

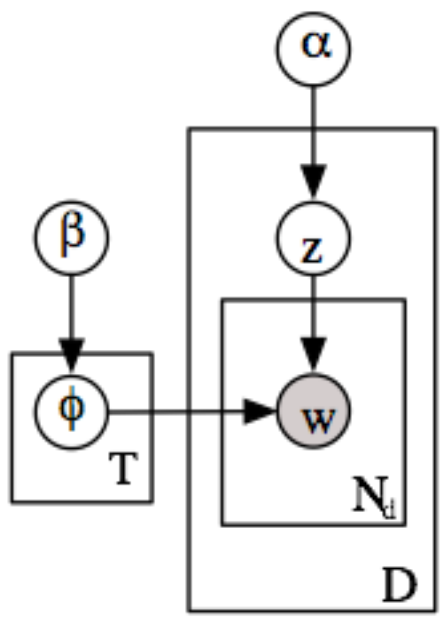
Introduction to Natural Language Processing
Computer Science 585—Fall 2009
University of Massachusetts Amherst

David Smith

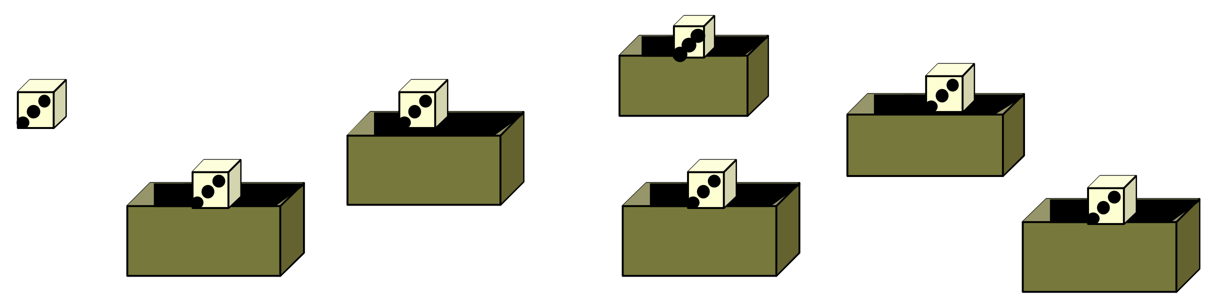
# Topic Models

Unsupervised Models of
Word Co-occurrences

# A Probabilistic Approach

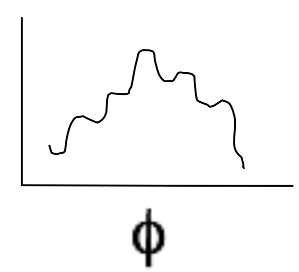



- Define a probabilistic generative model for documents.



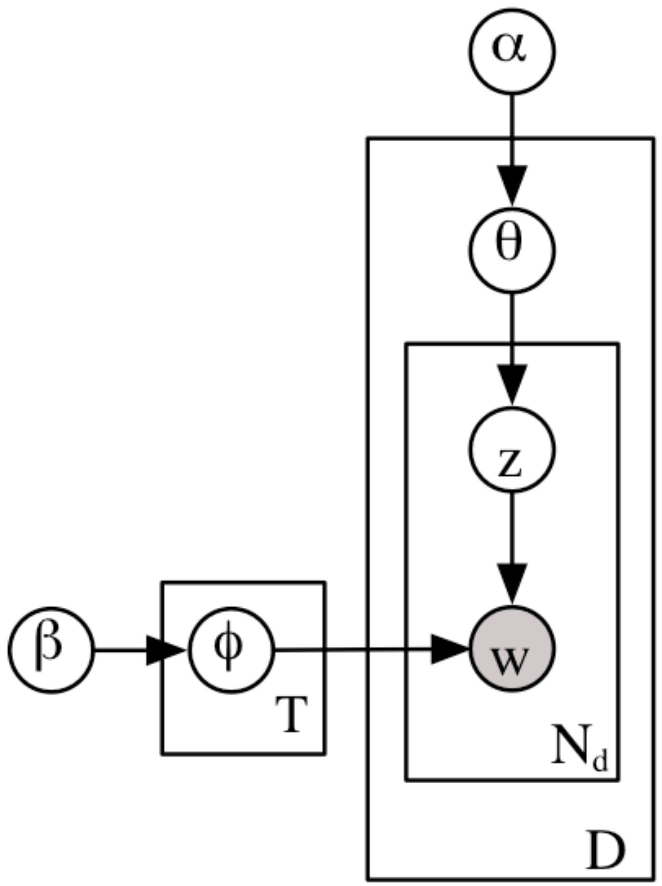
- Learn the parameters of this model by fitting them to the data and a prior.

$$\phi^* = \arg\max_\phi p(\phi|D_1 D_2 \ldots) = p(D_1 D_2 \ldots|\phi)\, p(\phi)$$



# Clustering words into topics with Latent Dirichlet Allocation

[Blei, Ng, Jordan 2003]




**Generative Process:**

*For each document:*

Sample a distribution over topics, $\theta$

*For each word in doc*

Sample a topic, $z$

Sample a word from the topic, $w$

**Example:**

70% Iraq war
30% US election

Iraq war

"bombing"

# Example topics
# induced from a large collection of text

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| DISEASE | WATER | MIND | STORY | FIELD | SCIENCE | BALL | JOB |
| BACTERIA | FISH | WORLD | STORIES | MAGNETIC | STUDY | GAME | WORK |
| DISEASES | SEA | DREAM | TELL | MAGNET | SCIENTISTS | TEAM | JOBS |
| GERMS | SWIM | DREAMS | CHARACTER | WIRE | SCIENTIFIC | FOOTBALL | CAREER |
| FEVER | SWIMMING | THOUGHT | CHARACTERS | NEEDLE | KNOWLEDGE | BASEBALL | EXPERIENCE |
| CAUSE | POOL | IMAGINATION | AUTHOR | CURRENT | WORK | PLAYERS | EMPLOYMENT |
| CAUSED | LIKE | MOMENT | READ | COIL | RESEARCH | PLAY | OPPORTUNITIES |
| SPREAD | SHELL | THOUGHTS | TOLD | POLES | CHEMISTRY | FIELD | WORKING |
| VIRUSES | SHARK | OWN | SETTING | IRON | TECHNOLOGY | PLAYER | TRAINING |
| INFECTION | TANK | REAL | TALES | COMPASS | MANY | BASKETBALL | SKILLS |
| VIRUS | SHELLS | LIFE | PLOT | LINES | MATHEMATICS | COACH | CAREERS |
| MICROORGANISMS | SHARKS | IMAGINE | TELLING | CORE | BIOLOGY | PLAYED | POSITIONS |
| PERSON | DIVING | SENSE | SHORT | ELECTRIC | FIELD | PLAYING | FIND |
| INFECTIOUS | DOLPHINS | CONSCIOUSNESS | FICTION | DIRECTION | PHYSICS | HIT | POSITION |
| COMMON | SWAM | STRANGE | ACTION | FORCE | LABORATORY | TENNIS | FIELD |
| CAUSING | LONG | FEELING | TRUE | MAGNETS | STUDIES | TEAMS | OCCUPATIONS |
| SMALLPOX | SEAL | WHOLE | EVENTS | BE | WORLD | GAMES | REQUIRE |
| BODY | DIVE | BEING | TELLS | MAGNETISM | SCIENTIST | SPORTS | OPPORTUNITY |
| INFECTIONS | DOLPHIN | MIGHT | TALE | POLE | STUDYING | BAT | EARN |
| CERTAIN | UNDERWATER | HOPE | NOVEL | INDUCED | SCIENCES | TERRY | ABLE |

**[Tennenbaum et al]**

# Example topics
# induced from a large collection of text

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| DISEASE | WATER | MIND | STORY | FIELD | SCIENCE | BALL | JOB |
| BACTERIA | FISH | WORLD | STORIES | MAGNETIC | STUDY | GAME | WORK |
| DISEASES | SEA | DREAM | TELL | MAGNET | SCIENTISTS | TEAM | JOBS |
| GERMS | SWIM | DREAMS | CHARACTER | WIRE | SCIENTIFIC | FOOTBALL | CAREER |
| FEVER | SWIMMING | THOUGHT | CHARACTERS | NEEDLE | KNOWLEDGE | BASEBALL | EXPERIENCE |
| CAUSE | POOL | IMAGINATION | AUTHOR | CURRENT | WORK | PLAYERS | EMPLOYMENT |
| CAUSED | LIKE | MOMENT | READ | COIL | RESEARCH | PLAY | OPPORTUNITIES |
| SPREAD | SHELL | THOUGHTS | TOLD | POLES | CHEMISTRY | FIELD | WORKING |
| VIRUSES | SHARK | OWN | SETTING | IRON | TECHNOLOGY | PLAYER | TRAINING |
| INFECTION | TANK | REAL | TALES | COMPASS | MANY | BASKETBALL | SKILLS |
| VIRUS | SHELLS | LIFE | PLOT | LINES | MATHEMATICS | COACH | CAREERS |
| MICROORGANISMS | SHARKS | IMAGINE | TELLING | CORE | BIOLOGY | PLAYED | POSITIONS |
| PERSON | DIVING | SENSE | SHORT | ELECTRIC | FIELD | PLAYING | FIND |
| INFECTIOUS | DOLPHINS | CONSCIOUSNESS | FICTION | DIRECTION | PHYSICS | HIT | POSITION |
| COMMON | SWAM | STRANGE | ACTION | FORCE | LABORATORY | TENNIS | FIELD |
| CAUSING | LONG | FEELING | TRUE | MAGNETS | STUDIES | TEAMS | OCCUPATIONS |
| SMALLPOX | SEAL | WHOLE | EVENTS | BE | WORLD | GAMES | REQUIRE |
| BODY | DIVE | BEING | TELLS | MAGNETISM | SCIENTIST | SPORTS | OPPORTUNITY |
| INFECTIONS | DOLPHIN | MIGHT | TALE | POLE | STUDYING | BAT | EARN |
| CERTAIN | UNDERWATER | HOPE | NOVEL | INDUCED | SCIENCES | TERRY | ABLE |

**[Tennenbaum et al]**

# Collocations

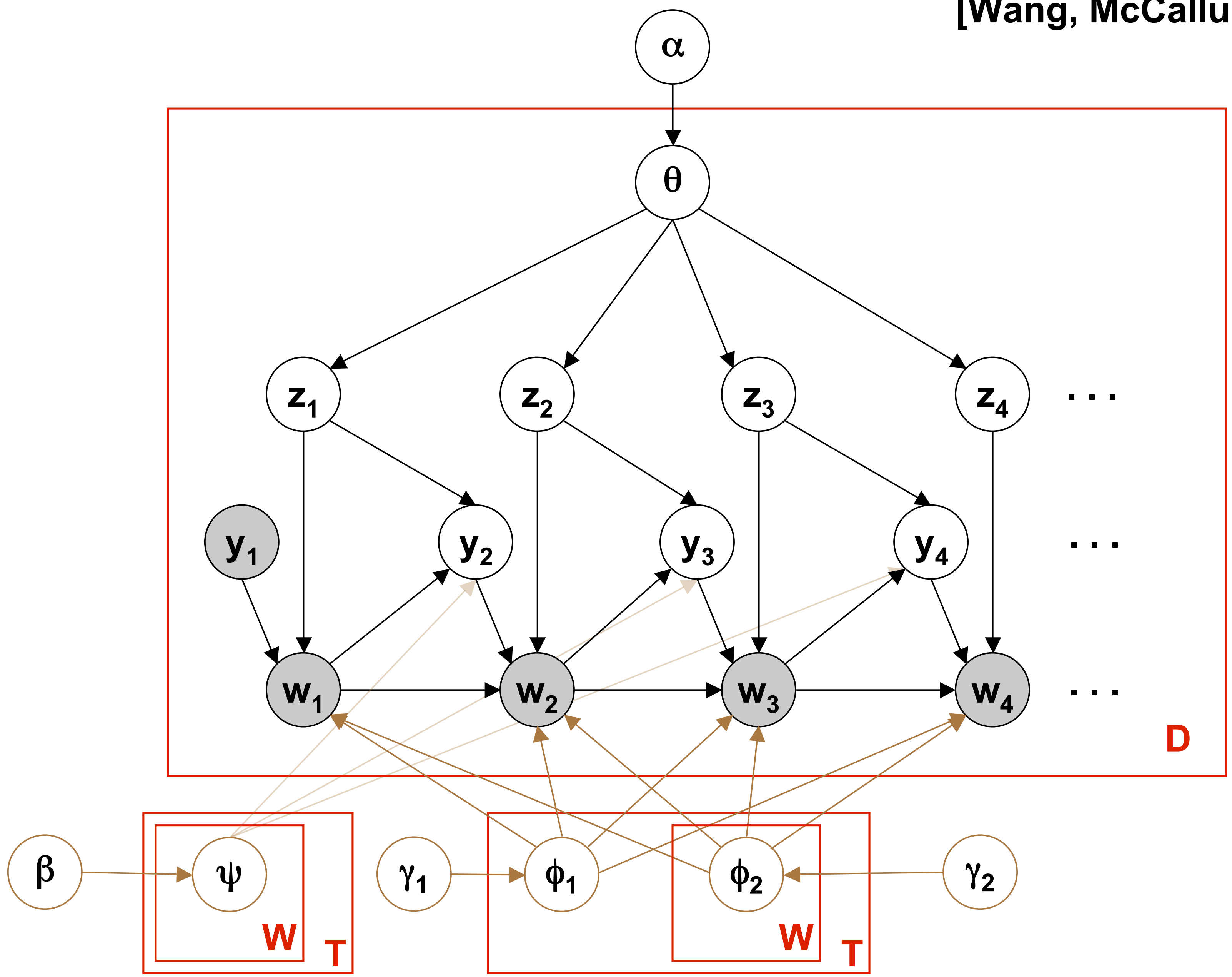
- An expression consisting of two or more words that correspond to some conventional way of saying things.
- Characterized by limited *compositionality*.
  - *compositional*: meaning of expression can be predicted by meaning of its parts.
  - “dynamic programming”, “hidden Markov model”
  - “weapons of mass destruction”
  - “kick the bucket”, “hear it through the grapevine”

## Topics Modeling Phrases

- Topics based only on unigrams often difficult to interpret
- Topic discovery itself is confused because important meaning / distinctions carried by phrases.
- Significant opportunity to provide improved language models to ASR, MT, IR, etc.

# Topical N-gram Model

[Wang, McCallum 2005]



## LDA Topic

| LDA | Topical N-grams |
|---|---|
| algorithms | genetic algorithms |
| algorithm | genetic algorithm |
| genetic | evolutionary computation |
| problems | evolutionary algorithms |
| efficient | fitness function |

# Topic Comparison

| LDA | Topical N-grams (2) | Topical N-grams (1) |
|---|---|---|
| learning | reinforcement learning | policy |
| optimal | optimal policy | action |
| reinforcement | dynamic programming | states |
| state | optimal control | actions |
| problems | function approximator | function |
| policy | prioritized sweeping | reward |
| dynamic | finite-state controller | control |
| action | learning system | agent |
| programming | reinforcement learning rl | q-learning |
| actions | function approximators | optimal |
| function | markov decision problems | goal |
| markov | markov decision processes | learning |
| methods | local search | space |
| decision | state-action pair | step |
| rl | markov decision process | environment |
| continuous | belief states | system |
| spaces | stochastic policy | problem |
| step | action selection | steps |
| policies | upright position | sutton |
| planning | reinforcement learning methods | policies |

# Topic Comparison

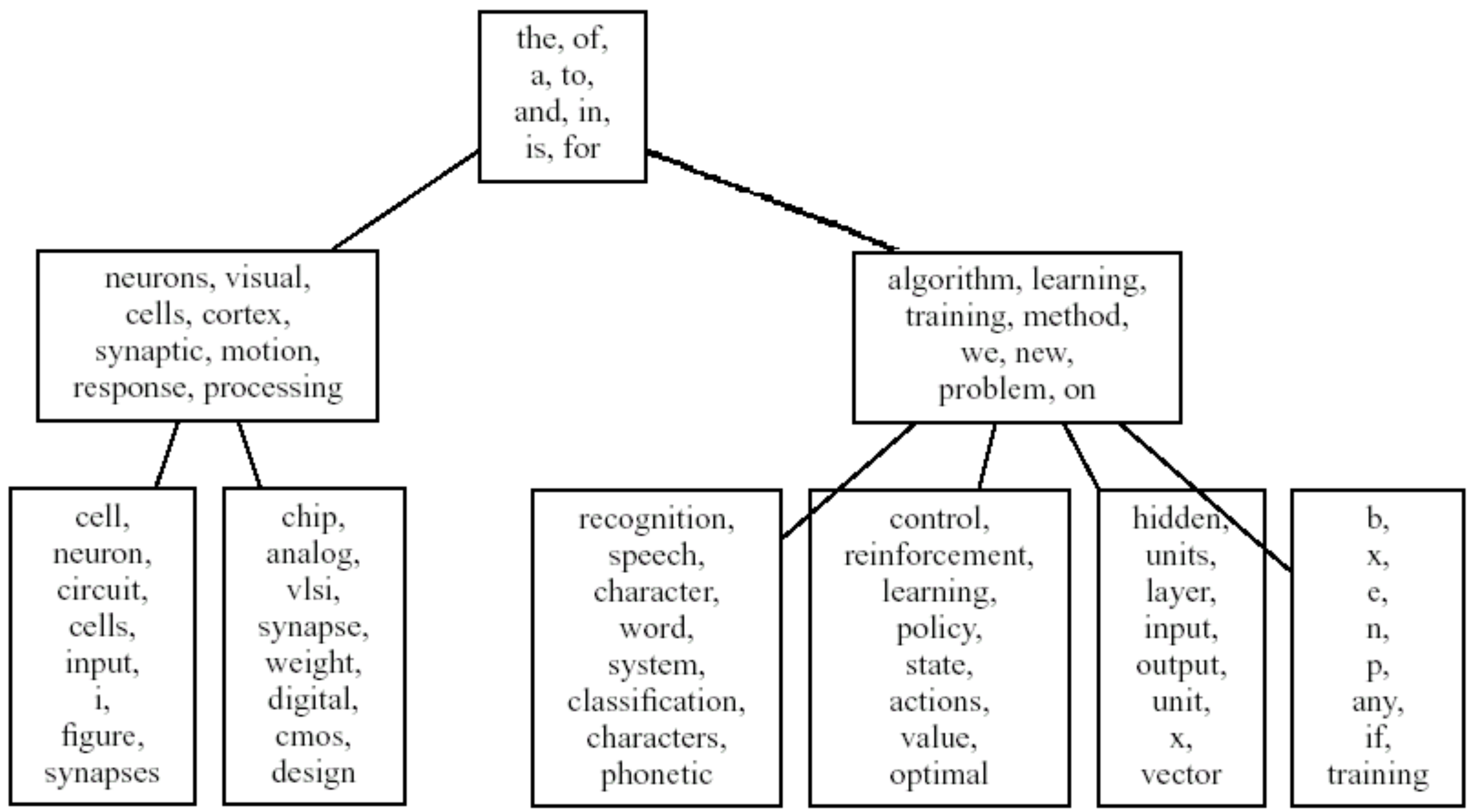
| LDA | Topical N-grams (2) | Topical N-grams (1) |
| --- | --- | --- |
| motion | receptive field | motion |
| visual | spatial frequency | response |
| field | temporal frequency | direction |
| position | visual motion | cells |
| figure | motion energy | stimulus |
| direction | tuning curves | figure |
| fields | horizontal cells | contrast |
| eye | motion detection | velocity |
| location | preferred direction | model |
| retina | visual processing | responses |
| receptive | area mt | stimuli |
| velocity | visual cortex | moving |
| vision | light intensity | cell |
| moving | directional selectivity | intensity |
| system | high contrast | population |
| flow | motion detectors | image |
| edge | spatial phase | center |
| center | moving stimuli | tuning |
| light | decision strategy | complex |
| local | visual stimuli | directions |

# Topic Comparison

| LDA | Topical N-grams (2) | Topical N-grams (1) |
|---|---|---|
| word | speech recognition | speech |
| system | training data | word |
| recognition | neural network | training |
| hmm | error rates | system |
| speech | neural net | recognition |
| training | hidden markov model | hmm |
| performance | feature vectors | speaker |
| phoneme | continuous speech | performance |
| words | training procedure | phoneme |
| context | continuous speech recognition | acoustic |
| systems | gamma filter | words |
| frame | hidden control | context |
| trained | speech production | systems |
| speaker | neural nets | frame |
| sequence | input representation | trained |
| speakers | output layers | sequence |
| mlp | training algorithm | phonetic |
| frames | test set | speakers |
| segmentation | speech frames | mlp |
| models | speaker dependent | hybrid |

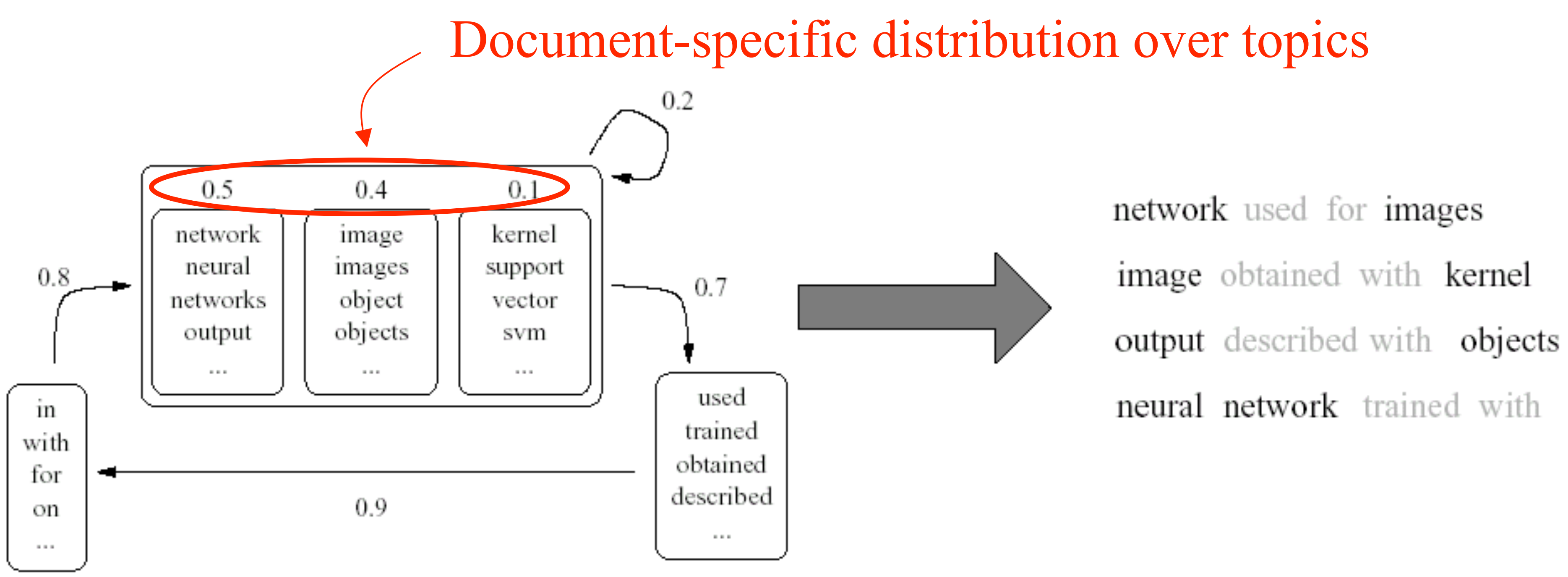
# Unsupervised learning of topic hierarchies

(Blei, Griffiths, Jordan & Tenenbaum, NIPS 2003)

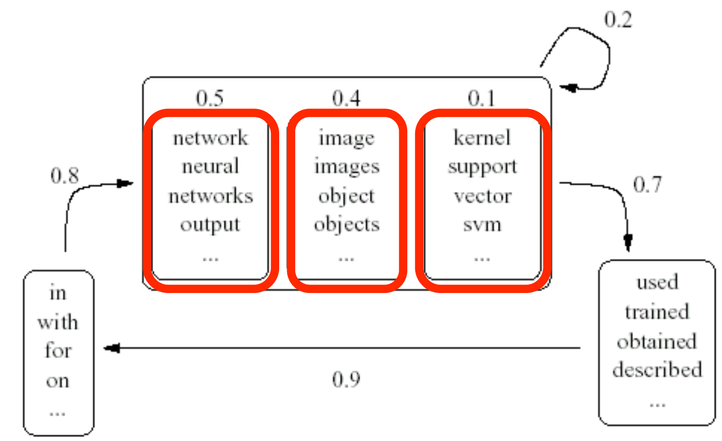


## Joint models of syntax and semantics (Griffiths, Steyvers, Blei & Tenenbaum, NIPS 2004)

- Embed topics model inside an *n*th order Hidden Markov Model:

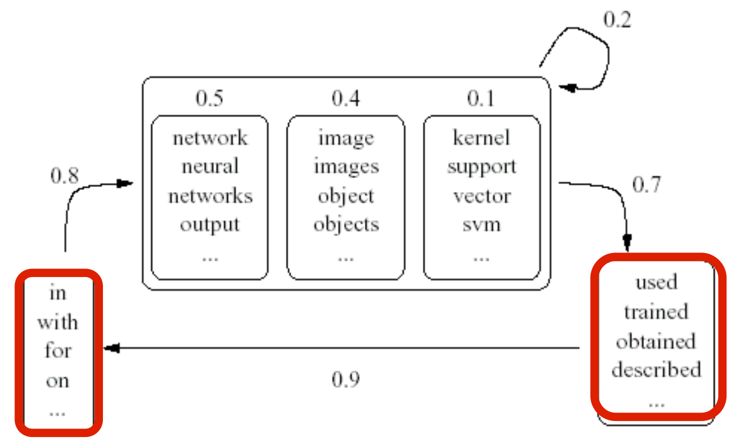


# Semantic classes




| | | | | | | | |
|---|---|---|---|---|---|---|---|
| FOOD | MAP | DOCTOR | BOOK | GOLD | BEHAVIOR | CELLS | PLANTS |
| FOODS | NORTH | PATIENT | BOOKS | IRON | SELF | CELL | PLANT |
| BODY | EARTH | HEALTH | READING | SILVER | INDIVIDUAL | ORGANISMS | LEAVES |
| NUTRIENTS | SOUTH | HOSPITAL | INFORMATION | COPPER | PERSONALITY | ALGAE | SEEDS |
| DIET | POLE | MEDICAL | LIBRARY | METAL | RESPONSE | BACTERIA | SOIL |
| FAT | MAPS | CARE | REPORT | METALS | SOCIAL | MICROSCOPE | ROOTS |
| SUGAR | EQUATOR | PATIENTS | PAGE | STEEL | EMOTIONAL | MEMBRANE | FLOWERS |
| ENERGY | WEST | NURSE | TITLE | CLAY | LEARNING | ORGANISM | WATER |
| MILK | LINES | DOCTORS | SUBJECT | LEAD | FEELINGS | FOOD | FOOD |
| EATING | EAST | MEDICINE | PAGES | ADAM | PSYCHOLOGISTS | LIVING | GREEN |
| FRUITS | AUSTRALIA | NURSING | GUIDE | ORE | INDIVIDUALS | FUNGI | SEED |
| VEGETABLES | GLOBE | TREATMENT | WORDS | ALUMINUM | PSYCHOLOGICAL | MOLD | STEMS |
| WEIGHT | POLES | NURSES | MATERIAL | MINERAL | EXPERIENCES | MATERIALS | FLOWER |
| FATS | HEMISPHERE | PHYSICIAN | ARTICLE | MINE | ENVIRONMENT | NUCLEUS | STEM |
| NEEDS | LATITUDE | HOSPITALS | ARTICLES | STONE | HUMAN | CELLED | LEAF |
| CARBOHYDRATES | PLACES | DR | WORD | MINERALS | RESPONSES | STRUCTURES | ANIMALS |
| VITAMINS | LAND | SICK | FACTS | POT | BEHAVIORS | MATERIAL | ROOT |
| CALORIES | WORLD | ASSISTANT | AUTHOR | MINING | ATTITUDES | STRUCTURE | POLLEN |
| PROTEIN | COMPASS | EMERGENCY | REFERENCE | MINERS | PSYCHOLOGY | GREEN | GROWING |
| MINERALS | CONTINENTS | PRACTICE | NOTE | TIN | PERSON | MOLDS | GROW |

# Syntactic classes




| | | | | | | | |
|---|---|---|---|---|---|---|---|
| SAID | THE | MORE | ON | GOOD | ONE | HE | BE |
| ASKED | HIS | SUCH | AT | SMALL | SOME | YOU | MAKE |
| THOUGHT | THEIR | LESS | INTO | NEW | MANY | THEY | GET |
| TOLD | YOUR | MUCH | FROM | IMPORTANT | TWO | I | HAVE |
| SAYS | HER | KNOWN | WITH | GREAT | EACH | SHE | GO |
| MEANS | ITS | JUST | THROUGH | LITTLE | ALL | WE | TAKE |
| CALLED | MY | BETTER | OVER | LARGE | MOST | IT | DO |
| CRIED | OUR | RATHER | AROUND | * | ANY | PEOPLE | FIND |
| SHOWS | THIS | GREATER | AGAINST | BIG | THREE | EVERYONE | USE |
| ANSWERED | THESE | HIGHER | ACROSS | LONG | THIS | OTHERS | SEE |
| TELLS | A | LARGER | UPON | HIGH | EVERY | SCIENTISTS | HELP |
| REPLIED | AN | LONGER | TOWARD | DIFFERENT | SEVERAL | SOMEONE | KEEP |
| SHOUTED | THAT | FASTER | UNDER | SPECIAL | FOUR | WHO | GIVE |
| EXPLAINED | NEW | EXACTLY | ALONG | OLD | FIVE | NOBODY | LOOK |
| LAUGHED | THOSE | SMALLER | NEAR | STRONG | BOTH | ONE | COME |
| MEANT | EACH | SOMETHING | BEHIND | YOUNG | TEN | SOMETHING | WORK |
| WROTE | MR | BIGGER | OFF | COMMON | SIX | ANYONE | MOVE |
| SHOWED | ANY | FEWER | ABOVE | WHITE | MUCH | EVERYBODY | LIVE |
| BELIEVED | MRS | LOWER | DOWN | SINGLE | TWENTY | SOME | EAT |
| WHISPERED | ALL | ALMOST | BEFORE | CERTAIN | EIGHT | THEN | BECOME |

# Corpus-specific factorization (NIPS)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Semantics | image | data | state | membrane | chip | experts | kernel | network |
| | images | gaussian | policy | synaptic | analog | expert | support | neural |
| | object | mixture | value | cell | neuron | gating | vector | networks |
| | objects | likelihood | function | * | digital | hme | svm | output |
| | feature | posterior | action | current | synapse | architecture | kernels | input |
| | recognition | prior | reinforcement | dendritic | neural | mixture | # | training |
| | views | distribution | learning | potential | hardware | learning | space | inputs |
| | # | em | classes | neuron | weight | mixtures | function | weights |
| | pixel | bayesian | optimal | conductance | # | function | machines | # |
| | visual | parameters | * | channels | vlsi | gate | set | outputs |
| Syntax | in | is | see | used | model | networks | however | # |
| | with | was | show | trained | algorithm | values | also | * |
| | for | has | note | obtained | system | results | then | i |
| | on | becomes | consider | described | case | models | thus | x |
| | from | denotes | assume | given | problem | parameters | therefore | t |
| | at | being | present | found | network | units | first | n |
| | using | remains | need | presented | method | data | here | - |
| | into | represents | propose | defined | approach | functions | now | c |
| | over | exists | describe | generated | paper | problems | hence | r |
| | within | seems | suggest | shown | process | algorithms | finally | p |

# Syntactic classes in PNAS

| 5 | 8 | 14 | 25 | 26 | 30 | 33 |
|---|---|---|---|---|---|---|
| IN | ARE | THE | SUGGEST | LEVELS | RESULTS | BEEN |
| FOR | WERE | THIS | INDICATE | NUMBER | ANALYSIS | MAY |
| ON | WAS | ITS | SUGGESTING | LEVEL | DATA | CAN |
| BETWEEN | IS | THEIR | SUGGESTS | RATE | STUDIES | COULD |
| DURING | WHEN | AN | SHOWED | TIME | STUDY | WELL |
| AMONG | REMAIN | EACH | REVEALED | CONCENTRATIONS | FINDINGS | DID |
| FROM | REMAINS | ONE | SHOW | VARIETY | EXPERIMENTS | DOES |
| UNDER | REMAINED | ANY | DEMONSTRATE | RANGE | OBSERVATIONS | DO |
| WITHIN | PREVIOUSLY | INCREASED | INDICATING | CONCENTRATION | HYPOTHESIS | MIGHT |
| THROUGHOUT | BECOME | EXOGENOUS | PROVIDE | DOSE | ANALYSES | SHOULD |
| THROUGH | BECAME | OUR | SUPPORT | FAMILY | ASSAYS | WILL |
| TOWARD | BEING | RECOMBINANT | INDICATES | SET | POSSIBILITY | WOULD |
| INTO | BUT | ENDOGENOUS | PROVIDES | FREQUENCY | MICROSCOPY | MUST |
| AT | GIVE | TOTAL | INDICATED | SERIES | PAPER | CANNOT |
| INVOLVING | MERE | PURIFIED | DEMONSTRATED | AMOUNTS | WORK | REMAINED |
| AFTER | APPEARED | TILE | SHOWS | RATES | EVIDENCE | ALSO |
| ACROSS | APPEAR | FULL | SO | CLASS | FINDING | THEY |
| AGAINST | ALLOWED | CHRONIC | REVEAL | VALUES | MUTAGENESIS | BECOME |
| WHEN | NORMALLY | ANOTHER | DEMONSTRATES | AMOUNT | OBSERVATION | MAG |
| ALONG | EACH | EXCESS | SUGGESTED | SITES | MEASUREMENTS | LIKELY |

# Semantic highlighting

Darker words are more likely to have been generated from the topic-based "semantics" module:

In contrast to this approach, we study here how the overall network activity can control single cell parameters such as input resistance, as well as time and space constants, parameters that are crucial for excitability and spariotemporal (sic) integration.

The integrated architecture in this paper combines feed forward control and error feedback adaptive control using neural networks.

---

In other words, for our proof of convergence, we require the softassign algorithm to return a doubly stochastic matrix as *sinkhorn theorem guarantees that it will instead of a matrix which is merely close to being doubly stochastic based on some reasonable metric.

The aim is to construct a portfolio with a maximal expected return for a given risk level and time horizon while simultaneously obeying *institutional or *legally required constraints.

---

The left graph is the standard experiment the right from a training with # samples.

The graph $G$ is called the *guest graph, and $H$ is called the host graph.

# PP Attachment:<br>A Simple Application of<br>Word Association

# Attachment Ambiguity

- Where to attach a phrase in the parse tree?
- *"I saw the man with the telescope."*
  - What does "with a telescope" modify?
  - Is the problem AI complete? Yes, but…
  - Proposed simple structural factors
    - Right association [Kimball 1973]
      'low' or 'near' attachment = 'early closure' of NP
    - Minimal attachment [Frazier 1978]
      (depends on grammar) = 'high' or 'distant' attachment
      = 'late closure' (of NP)

# Attachment Ambiguity

- “The children ate the cake with a spoon.”
- “The children ate the cake with frosting.”

- “Joe included the package for Susan.”
- “Joe carried the package for Susan.”

- *Ford, Bresnan and Kaplan (1982):*
  *“It is quite evident, then, that the closure effects in these sentences are induced in some way by the choice of the lexical items.”*

## Lexical acquisition, semantic similarity

- Previous models give same estimate to all unseen events.
- Unrealistic - could hope to refine that based on semantic classes of words
- Examples
  - “Susan ate the cake with a durian.”
  - “Susan had never eaten a fresh durian before.”
  - Although never seen “eating pineapple” should be more likely than “eating holograms” because pineapple is similar to apples, and we have seen “eating apples”.

# An application: selectional preferences

- Most verbs prefer arguments of a particular type. Such regularities are called *selectional preferences* or *selectional restrictions*.
- “Bill drove a…” Mustang, car, truck, jeep

- Selectional preference strength: how strongly does a verb constrain direct objects
- “see” versus “unknotted”

# Measuring selectional preference strength

- Assume we are given a clustering of (direct object) nouns. Resnick (1993) uses WordNet.

- Selectional association between a verb and a class

$$S(v) = D(P(C|v)||P(C) = \sum_{c} P(c|v) \log \frac{P(c|v)}{P(c)}$$

Proportion that its summand contributes to preference strength.

$$A(v,c) = \frac{P(c|v) \log \frac{P(c|v)}{P(c)}}{S(v)}$$

- For nouns in multiple classes, disambiguate as most likely sense: $A(v,n) = \max_{c \in \text{classes}(n)} A(v,c)$

# Selection preference strength (made up data)

| Noun class c | P(c) | P(c\|eat) | P(c\|see) | P(c\|find) |
|---|---|---|---|---|
| people | 0.25 | 0.01 | 0.25 | 0.33 |
| furniture | 0.25 | 0.01 | 0.25 | 0.33 |
| food | 0.25 | 0.97 | 0.25 | 0.33 |
| action | 0.25 | 0.01 | 0.25 | 0.01 |
| **SPS S(v)** | | **1.76** | **0.00** | **0.35** |

A(eat, food) = 1.08

A(find, action) = -0.13

# Selectional Preference Strength example

## (Resnick, Brown corpus)

| Verb $v$ | Noun $n$ | $A(v,n)$ | Class | Noun $n$ | $A(v,n)$ | Class |
|---|---|---|---|---|---|---|
| *answer* | *request* | 4.49 | speech act | *tragedy* | 3.88 | communication |
| *find* | *label* | 1.10 | abstraction | *fever* | 0.22 | psych. feature |
| *hear* | *story* | 1.89 | communication | *issue* | 1.89 | communication |
| *remember* | *reply* | 1.31 | statement | *smoke* | 0.20 | article of commerce |
| *repeat* | *comment* | 1.23 | communication | *journal* | 1.23 | communication |
| *read* | *article* | 6.80 | writing | *fashion* | −0.20 | activity |
| *see* | *friend* | 5.79 | entity | *method* | −0.01 | method |
| *write* | *letter* | 7.26 | writing | *market* | 0.00 | commerce |

# But how might we measure word similarity for word classes?

- Vector spaces

A document-by-word matrix $A$.

| | cosmonaut | astronaut | moon | car | truck |
|---|---|---|---|---|---|
| $d_1$ | 1 | 0 | 1 | 1 | 0 |
| $d_2$ | 0 | 1 | 1 | 0 | 0 |
| $d_3$ | 1 | 0 | 0 | 0 | 0 |
| $d_4$ | 0 | 0 | 0 | 1 | 1 |
| $d_5$ | 0 | 0 | 0 | 1 | 0 |
| $d_6$ | 0 | 0 | 0 | 0 | 1 |

# But how might we measure word similarity for word classes?

- Vector spaces

**word-by-word matrix B**

| | cosmonaut | astronaut | moon | car | truck |
|---|---|---|---|---|---|
| cosmonaut | 2 | 0 | 1 | 1 | 0 |
| astronaut | 0 | 1 | 1 | 0 | 0 |
| moon | 1 | 1 | 2 | 1 | 0 |
| car | 1 | 0 | 1 | 3 | 1 |
| truck | 0 | 0 | 0 | 1 | 2 |

**A modifier-by-head matrix $C$**

| | cosmonaut | astronaut | moon | car | truck |
|---|---|---|---|---|---|
| Soviet | 1 | 0 | 0 | 1 | 1 |
| American | 0 | 1 | 0 | 1 | 1 |
| spacewalking | 1 | 1 | 0 | 0 | 0 |
| red | 0 | 0 | 0 | 1 | 1 |
| full | 0 | 0 | 1 | 0 | 0 |
| old | 0 | 0 | 0 | 1 | 1 |

# Similarity measures for binary vectors

| Similarity measure | Definition |
|---|---|
| matching coefficient | $\lvert X \cap Y \rvert$ |
| Dice coefficient | $\frac{2\lvert X \cap Y \rvert}{\lvert X \rvert + \lvert Y \rvert}$ |
| Jaccard coefficient | $\frac{\lvert X \cap Y \rvert}{\lvert X \cup Y \rvert}$ |
| Overlap coefficient | $\frac{\lvert X \cap Y \rvert}{\min(\lvert X \rvert, \lvert Y \rvert)}$ |
| cosine | $\frac{\lvert X \cap Y \rvert}{\sqrt{\lvert X \rvert \times \lvert Y \rvert}}$ |

## Cosine measure

$$\cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}||\vec{y}|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

maps vectors onto unit circle by dividing through

by lengths:

$$|\vec{x}| = \sqrt{\sum_{i=1}^{n} x_i^2}$$

# Example of cosine measure on word-by-word matrix on NYT

| Focus word | Nearest neighbors | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *garlic* | *sauce* | .732 | *pepper* | .728 | *salt* | .726 | *cup* | .726 |
| *fallen* | *fell* | .932 | *decline* | .931 | *rise* | .930 | *drop* | .929 |
| *engineered* | *genetically* | .758 | *drugs* | .688 | *research* | .687 | *drug* | .685 |
| *Alfred* | *named* | .814 | *Robert* | .809 | *William* | .808 | *W* | .808 |
| *simple* | *something* | .964 | *things* | .963 | *You* | .963 | *always* | .962 |

# Probabilistic measures

| (Dis-)similarity measure | Definition |
|---|---|
| KL divergence | $D(p\|q) = \sum_i p_i \log \frac{p_i}{q_i}$ |
| Skew | $D(q\|\alpha r + (1-\alpha)q)$ |
| Jensen-Shannon (was IRad) | $\frac{1}{2}D(p\|\frac{p+q}{2}) + D(q\|\frac{p+q}{2})$ |
| $L_1$ norm (Manhattan) | $\sum_i \lvert p_i - q_i \rvert$ |

# Neighbors of word "company"
## [Lee]

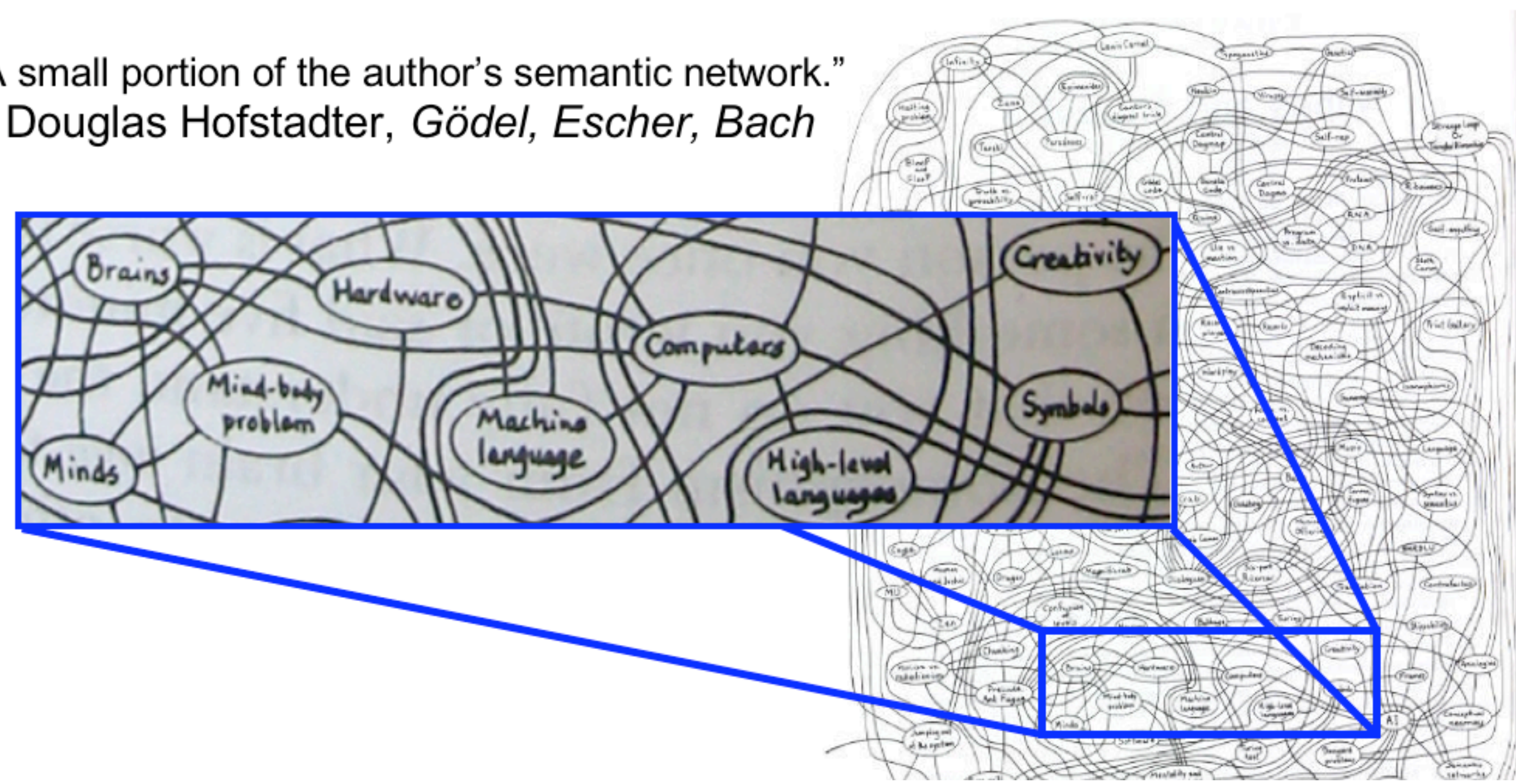
| Skew ($\alpha = 0.99$) | J.-S. | Euclidean |
|---|---|---|
| airline | business | city |
| business | airline | airline |
| bank | firm | industry |
| agency | bank | program |
| firm | state | organization |
| department | agency | bank |
| manufacturer | group | system |
| network | govt. | today |
| industry | city | series |
| govt. | industry | portion |

# Learning syntactic patterns for automatic hypernym discovery

Rion Snow, Daniel Jurafsky, and Andrew Y. Ng.

- It has long been a goal of AI to automatically acquire structured knowledge directly from text, e.g, in the form of a semantic network.

"A small portion of the author's semantic network."
– Douglas Hofstadter, *Gödel, Escher, Bach*



We aim to classify whether a noun pair $(X, Y)$ participates in one of the following semantic relationships:

**Hypernymy** (ancestor)

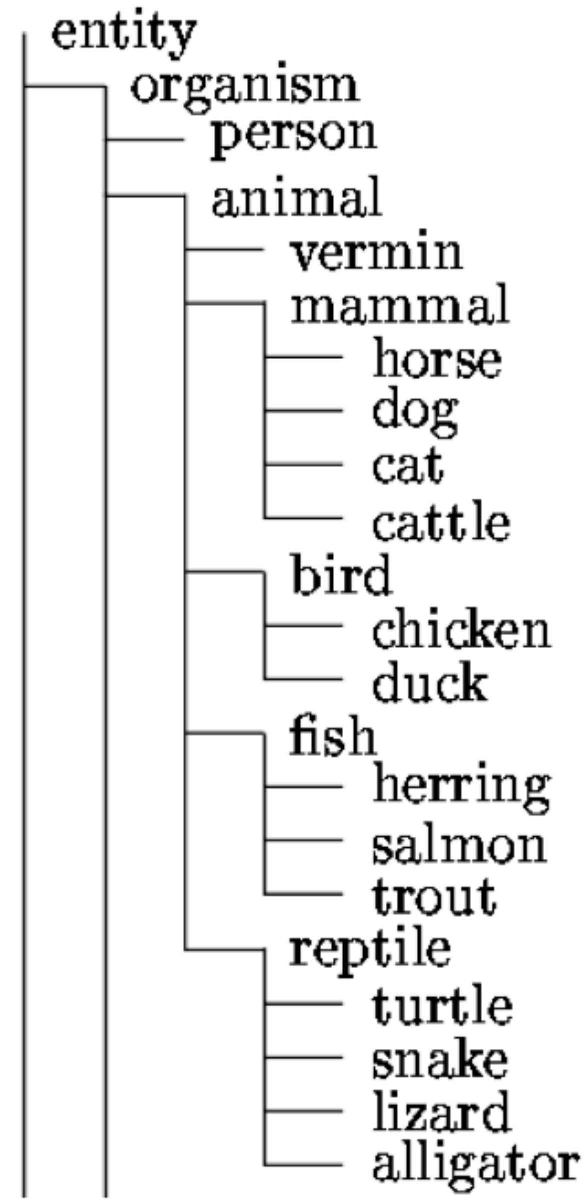
$Y \underset{H}{>} X$ if "$X$ is a kind of $Y$".

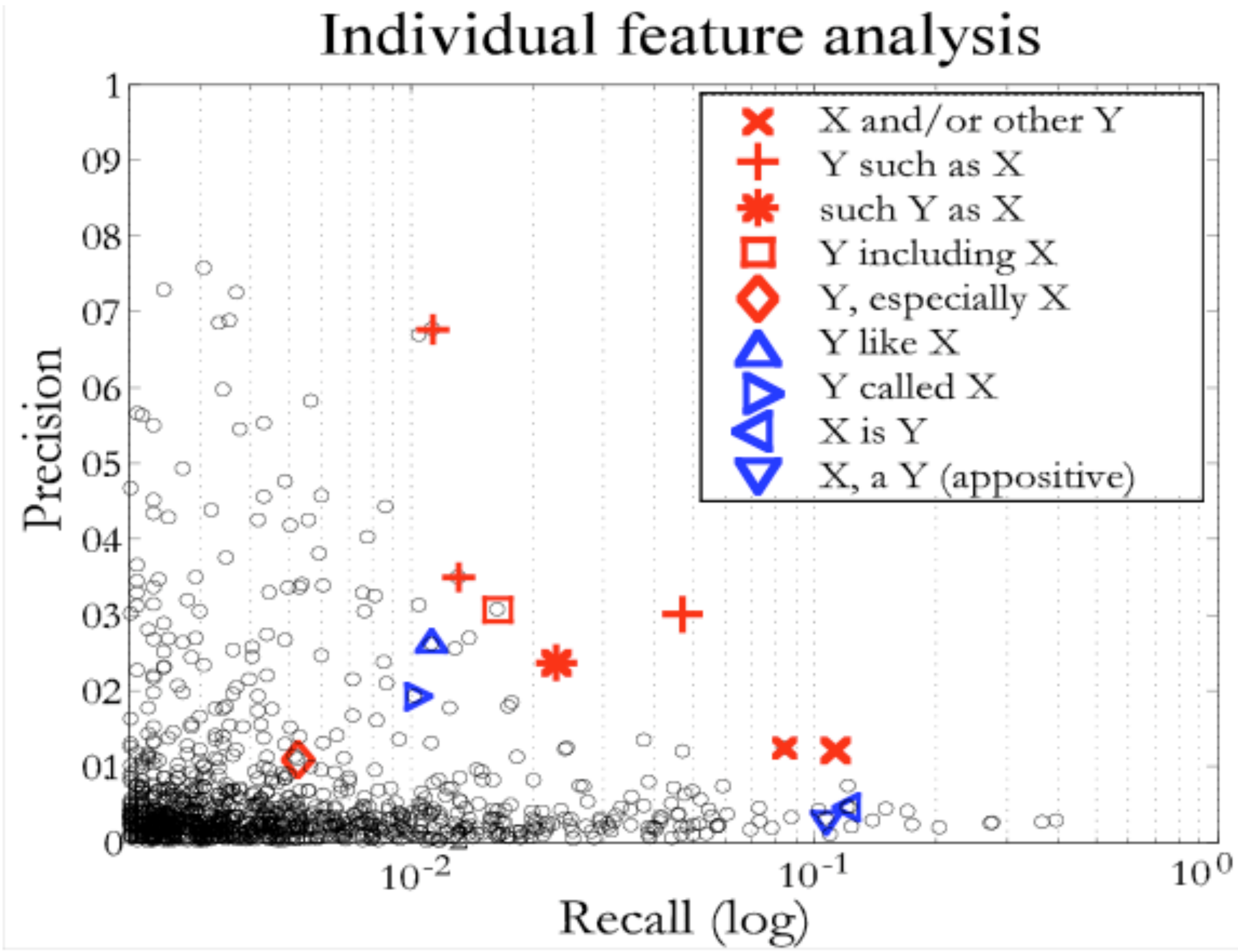
$$entity \underset{H}{>} organism \underset{H}{>} person$$

**Coordinate Terms** (taxonomic sisters)

$Y \underset{C}{\sim} X$ if $X$ and $Y$ possess a common hypernym, i.e. $\exists Z$ such that "$X$ and $Y$ are both kinds of $Z$."

$$horse \underset{C}{\sim} dog \underset{C}{\sim} cat$$

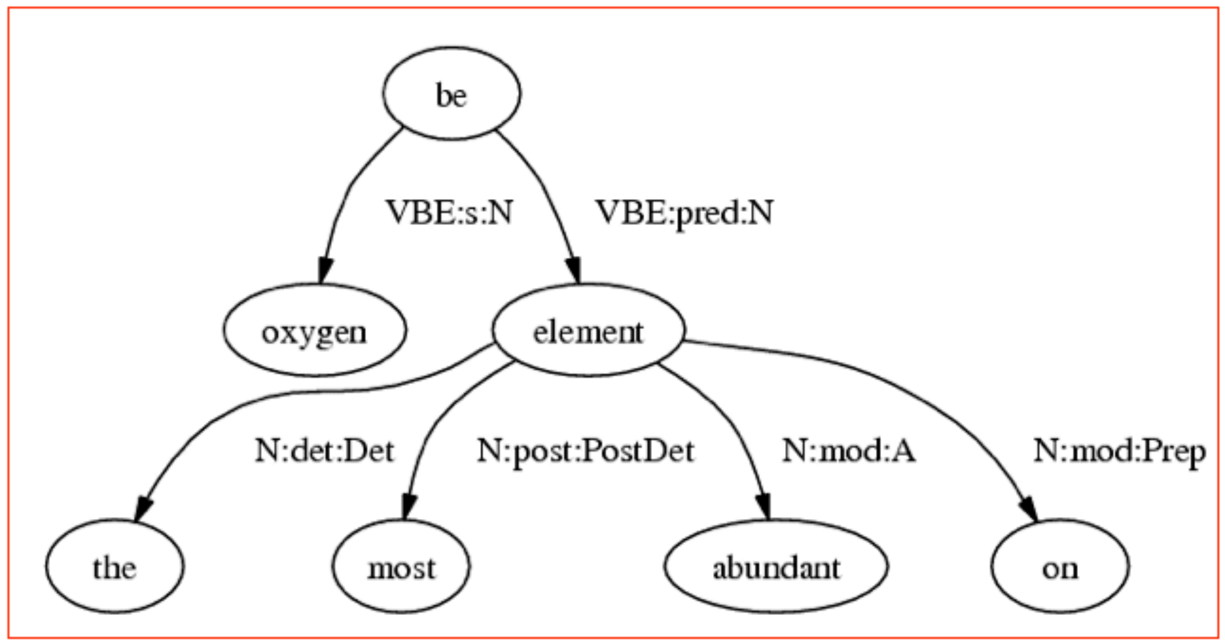





- Precision/recall for 69,592 classifiers (one per feature)
- Classifier $f$ classifies noun pair $\boldsymbol{x}$ as hypernym iff $x_f > 0$
- **In red:** patterns originally proposed in (Hearst, 1992)

**“Oxygen is the most abundant element on the moon.”**

## Dependency Graph:

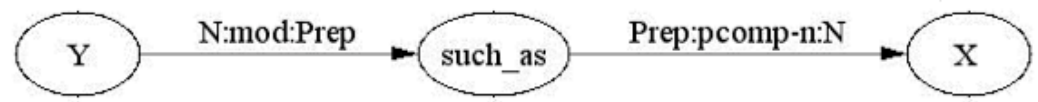



## Dependency Paths (for “oxygen / element” ):

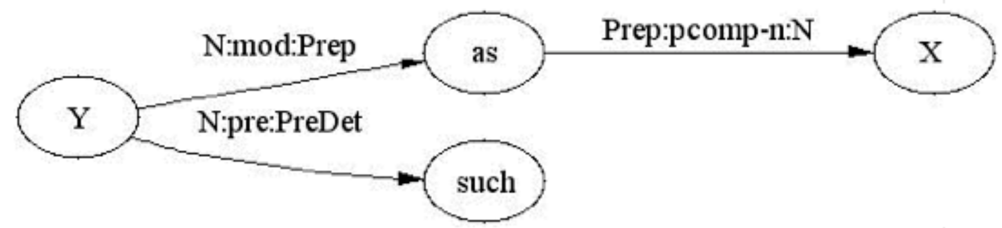
**-N:s:VBE, “be” VBE:pred:N**
**-N:s:VBE, “be” VBE:pred:N,(the,Det:det:N)**
**-N:s:VBE, “be” VBE:pred:N,(most,PostDet:post:N)**
**-N:s:VBE, “be” VBE:pred:N,(abundant,A:mod:N)**
**-N:s:VBE, “be” VBE:pred:N,(on,Prep:mod:N)**

# Rediscovering Hearst's Patterns

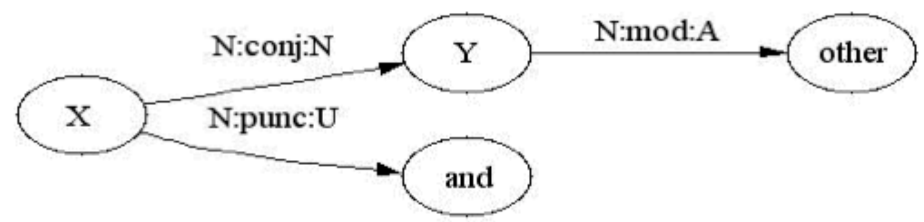
Y such as X…




Such Y as X…




X… and other Y




Proposed in (Hearst, 1992) and used in (Caraballo, 2001), (Widdows, 2003), and others – but what about the rest of the lexico-syntactic pattern space?

## Example: Using the "Y called X" Pattern for Hypernym Acquisition

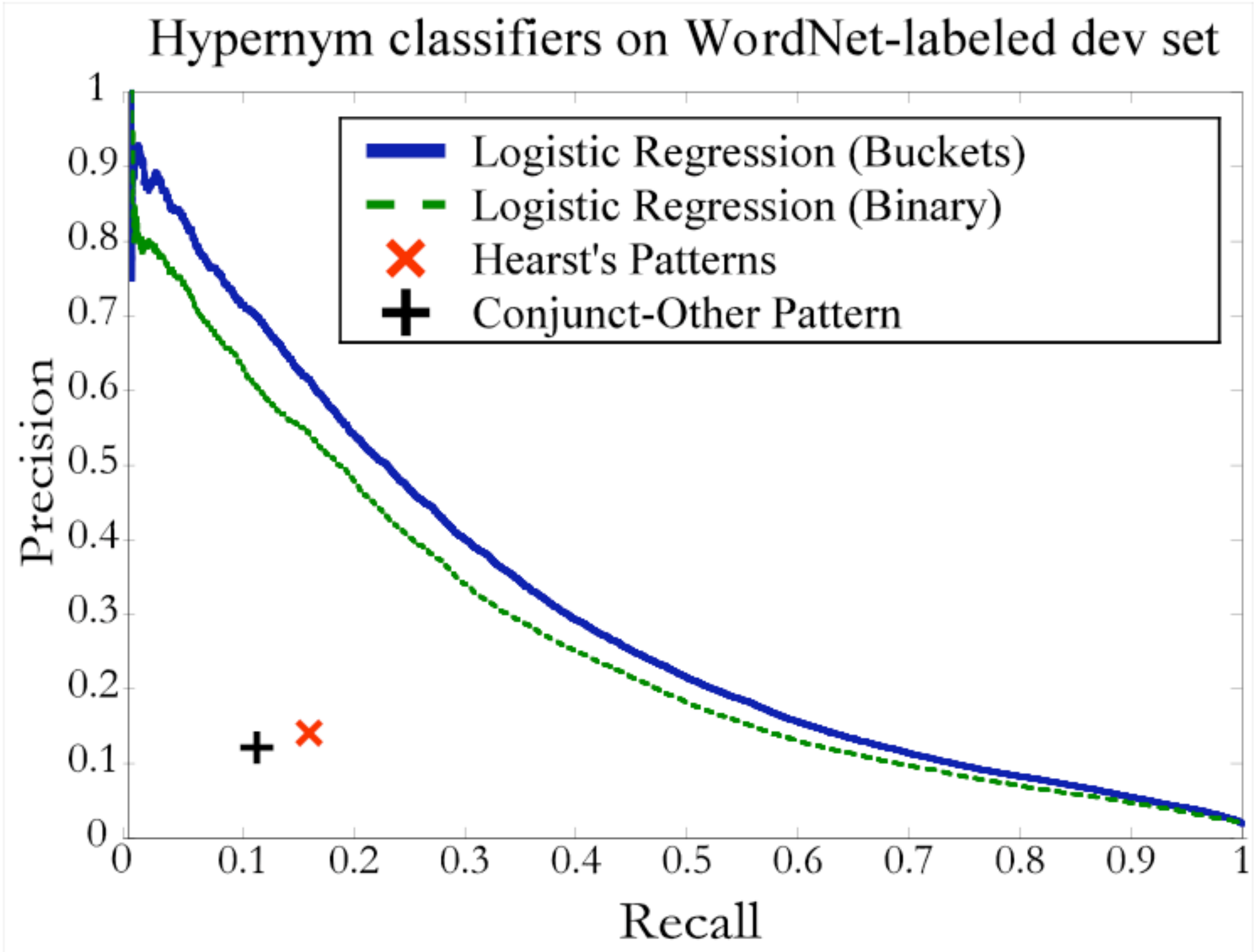
**MINIPAR path: -N:desc:V,call,call,-V:vrel:N → "<hypernym> 'called' <hyponym>"**

None of the following links are contained in WordNet (or the training set, by extension).

| Hyponym | Hypernym | Sentence Fragment |
|---|---|---|
| **efflorescence** | **condition** | …and a **condition called efflorescence**… |
| **'neal_inc** | **company** | …The **company**, now **called O'Neal Inc.**… |
| **hat_creek_outfit** | **ranch** | …run a small **ranch called** the **Hat Creek Outfit**. |
| **tardive_dyskinesia** | **problem** | ... irreversible **problem called tardive dyskinesia**… |
| **hiv-1** | **aids_virus** | …infected by the **AIDS virus**, **called HIV-1**. |
| **bateau_mouche** | **attraction** | …sightseeing **attraction called** the **Bateau Mouche**... |
| **kibbutz_malkiyya** | **collective_farm** | …Israeli **collective farm called Kibbutz Malkiyya**… |

| Type of Noun Pair | Count | Example Pair |
|---|---|---|
| NE: Person | 7 | "John F. Kennedy / president", "Marlin Fitzwater / spokesman" |
| NE: Place | 7 | "Diamond Bar / city", "France / place" |
| NE: Company | 2 | "American Can / company", "Simmons / company" |
| NE: Other | 1 | "Is Elvis Alive / book" |
| Not Named Entity: | 9 | "earthquake / disaster", "soybean / crop" |

# A better hypernym classifier




- 10-fold cross validation on the WordNet-labeled data
- **Conclusion:** 70,000 features are more powerful than 6

# VERBOCEAN: Mining the Web for Fine-Grained Semantic Verb Relations

Timothy Chklovski and Patrick Pantel

# Why Detect Semantic Rels between Verbs?

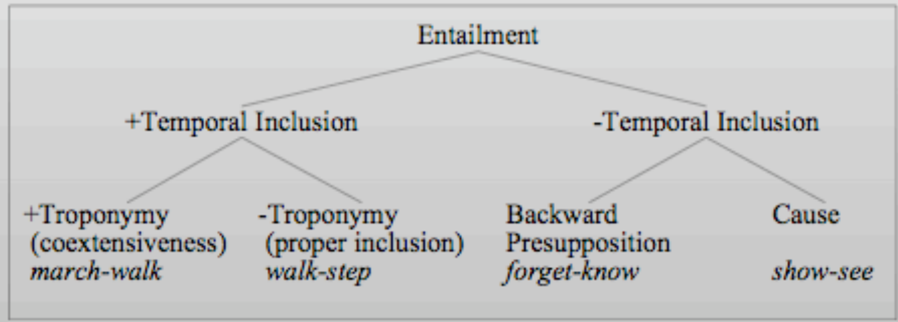
- So that we can
  - Understand the relationship when it's not stated
    - Napoleon ***fought*** and ***won*** the battle
    - During the holidays, people ***wrap*** and ***unwrap*** presents
    - Soldiers prefer to avoid getting ***wounded*** and ***killed***
  - Use the relationship when summarizing across documents (e.g. same event, preceding event)
    - The board ***considered*** the offer of $3B
    - The board ***accepted*** the offer $3.8B
    - The board ***okayed*** the offer of approximately $4B
  - Determine if two people have similar views on and event
    - "I ***nudged*** him."
    - "He ***shoved*** me."
- Hard to do manually

# Why use Web? Motivating Intuition

- Small collections are tough: Semantics is often implied (Lenat, Chklovski)
- The Web's $\mathbf{10^{12}}$ is a lot of words
- So, Use small bits of more detailed text to help with mass of general text
  - Patterns issued to a search engine and their correlation

# Relevant Work

- Levin's classes (similarity)
  - 3200 verbs in 191 classes
- PropBank
  - 4,659 framesets (1.4 framesets per verb)
- VerbNet
  - 191 coarse-grained groupings (with overlap)
- FrameNet
- WordNet
  - troponomy
  - antonymy
  - entailment
  - cause




Fellbaum's (1998) entailment hierarchy.

# VerbOcean: Web-based Extraction of Verb Relations

- VerbOcean is a network of verb relations
  - Currently, over 3400 nodes with on average 13 relations per verb
- Detected relation types are:
  - similarity
  - strength
  - antonymy
  - enablement
  - temporal precedence (happens-before)
- Download from http://semantics.isi.edu/ocean/

# Approach

- Three stages:
  - Identify pairs of highly associated verbs co-occurring on the Web with sufficient frequency using DIRT (Lin and Pantel 2001)
  - For each verb pair
    - test patterns associated with each semantic relation
      - E.g. Temporal Precedence:
        "to *X* and then *Y*", "*X*ed and then *Y*ed"
    - calculate a score for each possible semantic relation
  - Compare the strengths of the individual semantic relations and output a consistent set as the final output
    - prefer the most specific and then strongest relations

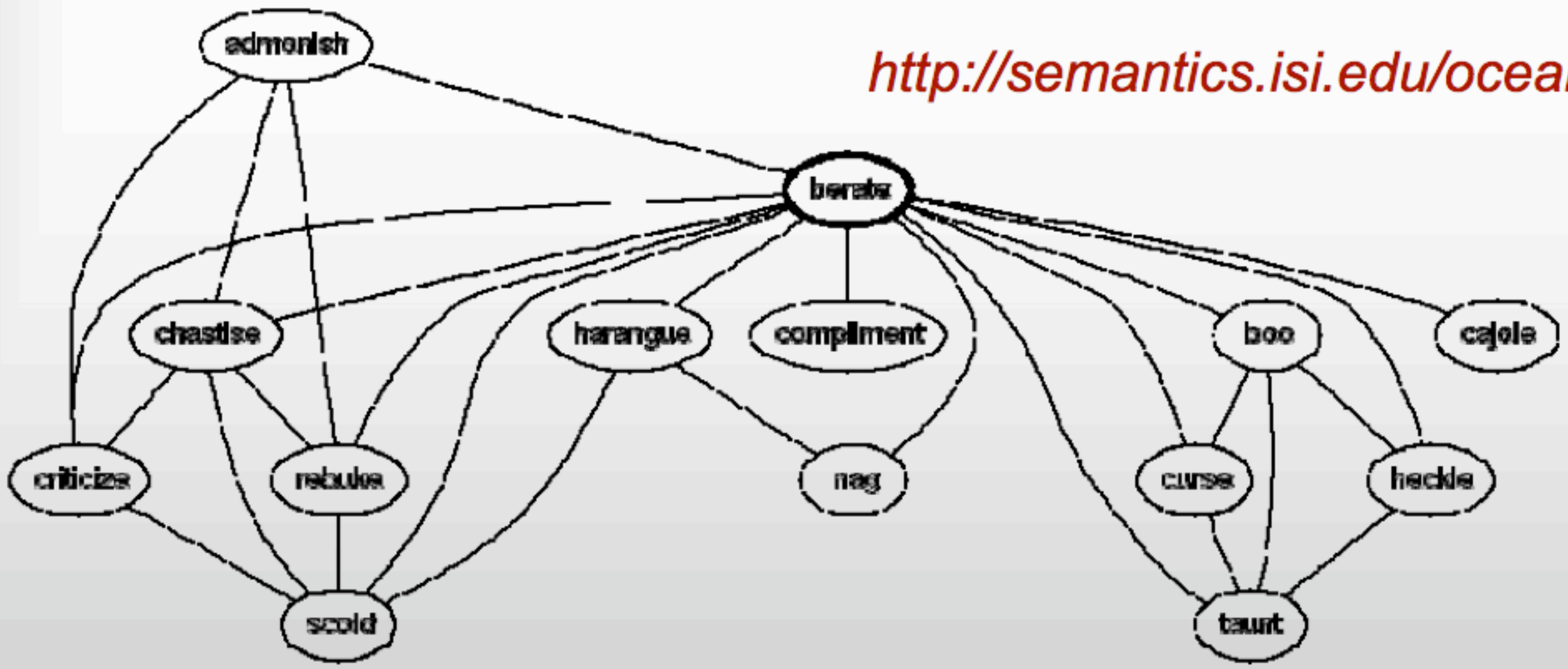
# Lexical Patterns

| *SEMANTIC RELATION* | *Surface Patterns* | *Example* |
|---|---|---|
| similarity (4) | X ie Y<br>Xed and Yed | *"She **heckled and taunted** the comedian."* |
| strength (8) | X even Y<br>Xed even Yed<br>Xed and even Yed<br>not just Xed but Yed | *"He not **just harassed, but terrorized** her."* |
| enablement (4) | Xed * by Ying the<br>Xed * by Ying or<br>to X * by Ying the | *"She **saved the document by clicking the** button."* |
| antonymy (7) | either X or Y<br>either Xs or Ys<br>Xed * but Yed | *"There's something about Mary: you will **either love or hate** her."* |
| happens-before (12) | to X and then Y<br>Xed * and then Yed<br>to X and later Y<br>to X and subsequently Y<br>Xed and subsequently Yed | *"He **designed the prototype and then patented** it."* |

# Lexical Patterns Match…

- Refined to decrease capturing wrong parts of speech or incorrect semantic relations
  - Xed * by Ying the; Xed * by Ying or
    - "… waved at by parking guard …"
    - "… encouraged further by sailing lessons …"

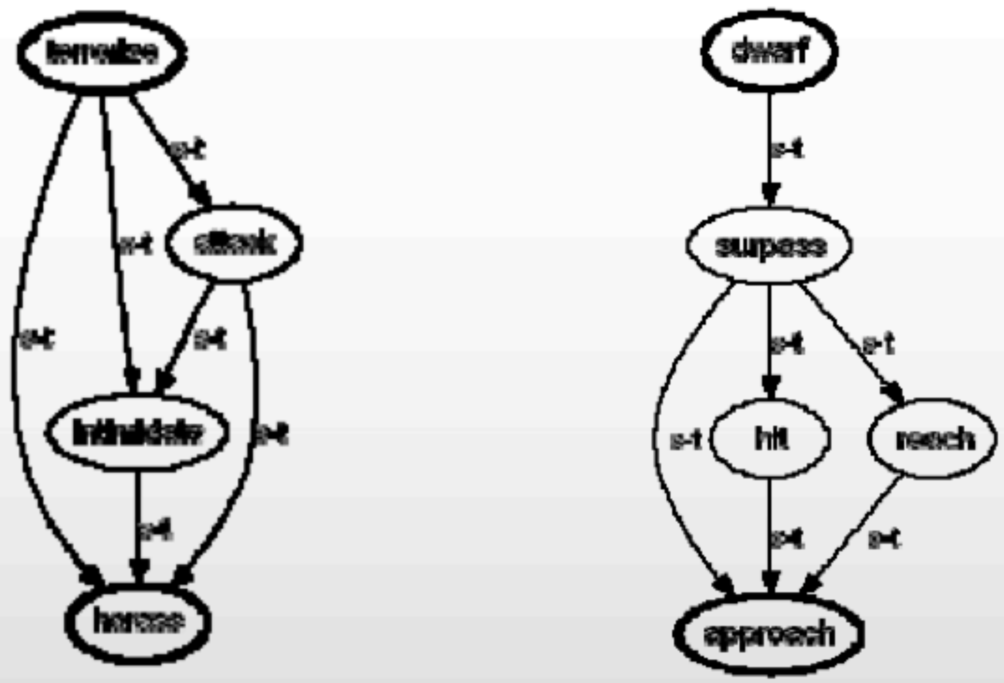
# VerbOcean – Similarity




*http://semantics.isi.edu/ocean/*

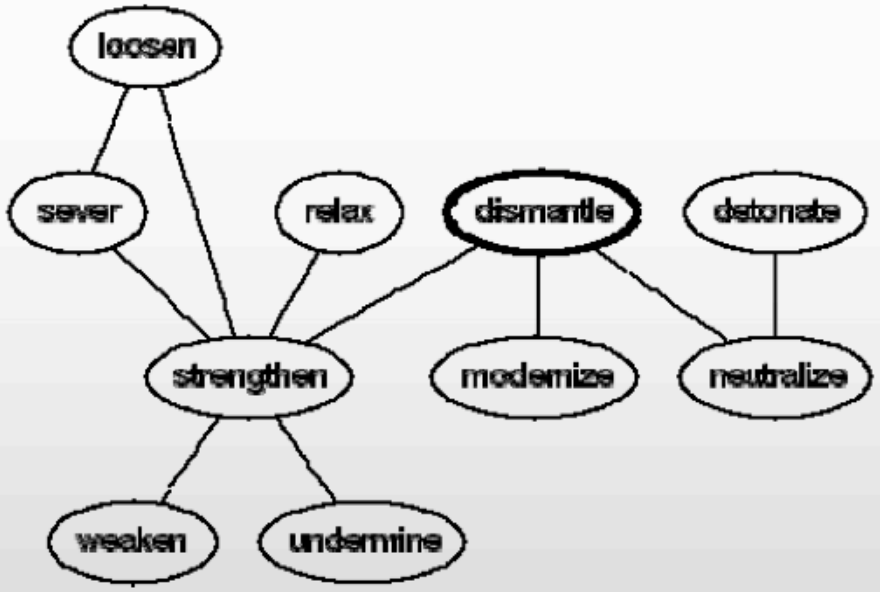
- Verbs that are similar or related
  - e.g. boo - heckle

# VerbOcean – Strength



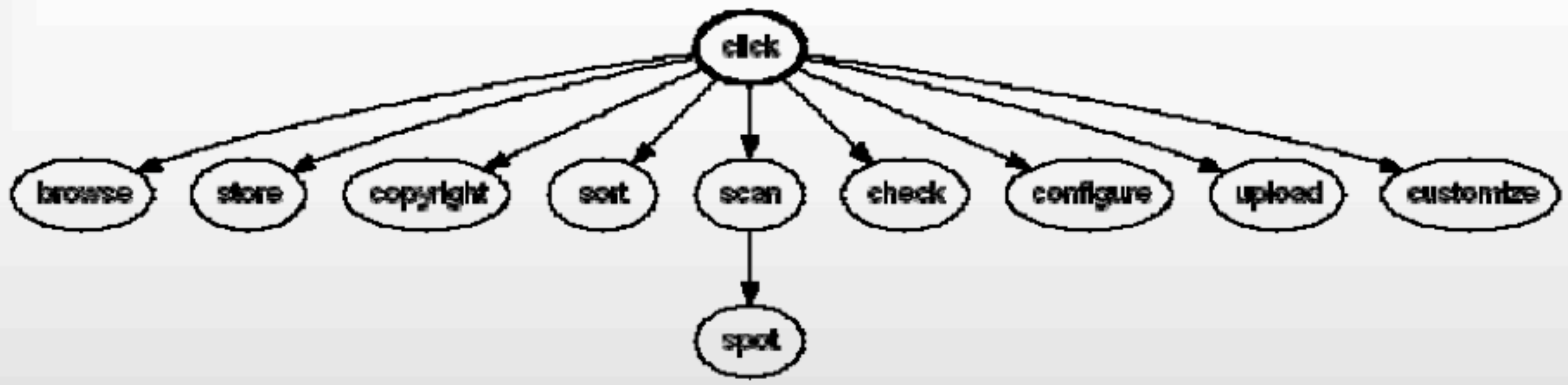

- Similar verbs that denote a more intense, thorough, comprehensive or absolute action
  - e.g. change-of-state verbs that denote a more complete change (shock → startle)

# VerbOcean – Antonymy




- Semantic opposition
  - switching thematic roles associated with the verb (buy – sell)
  - stative verbs (live – die)
  - sibling verbs which share a parent (walk – run)
  - restitutive opposition: antonymy + happens-before (damage - repair)

# VerbOcean – Enablement




- Holds between two verbs $V_1$ and $V_2$ when the pair can be glossed as "$V_1$ is accomplished by $V_2$" (assess - review)

**Appendix.** Sample relations extracted by our system.

| SEMANTIC RELATION | EXAMPLES | SEMANTIC RELATION | EXAMPLES | SEMANTIC RELATION | EXAMPLES |
|---|---|---|---|---|---|
| similarity | maximize :: enhance<br>produce :: create<br>reduce :: restrict | enablement | assess :: review<br>accomplish :: complete<br>double-click :: click | happens before | detain :: prosecute<br>enroll :: graduate<br>schedule :: reschedule |
| strength | permit :: authorize<br>surprise :: startle<br>startle :: shock | antonymy | assemble :: dismantle<br>regard :: condemn<br>roast :: fry | | |