Word Senses

Marianna Apidianaki
LIMSI, CNRS
marianna@limsi.fr

Computational Linguistics Course, UPenn
March, 28 2018

References

- Daniel Jurafsky and James H. Martin (2016 draft) Speech and Language Processing. *Chapter 17: Computing with Word Senses*.

Optional

- Peter D. Turney and Patrick Pantel (2010) From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research* (JAIR), Vol. 37, pages 141-188.
- Diana McCarthy, Marianna Apidianaki and Katrin Erk (2016) Word Sense Clustering and Clusterability. Computational Linguistics, Vol. 42(2), pp. 245-275.
- Marco Baroni, Georgiana Dinu and Germán Kruszewski (2014) *Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors*. Proceedings of ACL, Baltimore, Maryland, pages 238-247.

Overview

1. Word Senses and Relations
2. Word Sense Disambiguation
3. Word Sense Induction
4. Word Similarity

Part 1
Word Senses and Relations

Words can have different senses

bank

FINANCIAL INSTITUTION

SLOPING SIDE OF A RIVER

Words can have different senses

Homonymy

bank

a bank can hold the investments in a custodial account in the client's name

But as agriculture burgeons on the east bank, the river will shrink even more.

bank^1: FINANCIAL INSTITUTION

bank^2: SLOPING SIDE OF A RIVER

Homonymy vs Polysemy

Polysemy

a semantic connection exists between the senses

While some __banks__ furnish blood only to hospitals, others are less restrictive.

bank^3: BIOLOGICAL REPOSITORY

(*blood bank, egg bank, sperm bank, …*)

bank^1 ⟷ **bank^3**

Homonymy vs Polysemy

Systematic Polysemy

The relation between the senses is systematic and structured

The *bank* is on the corner of Nassau and Witherspoon.

bank^4: THE BUILDING BELONGING TO A FINANCIAL INSTITUTION

(school, university, hospital …)

BUILGING ↔ ORGANIZATION

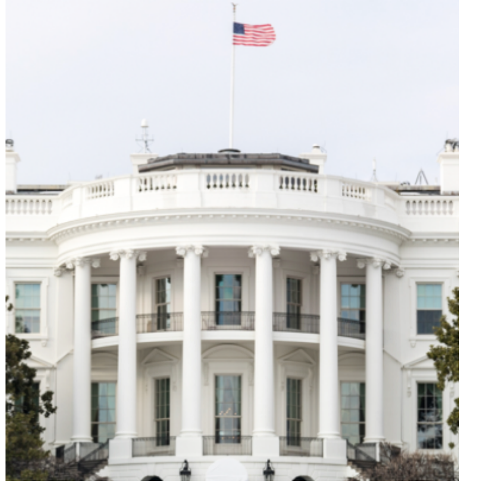
also called
Metonymy

Metonymy

Use of one aspect of a concept or entity to refer to other aspects of the entity or to the entity itself

> the <u>White House</u> announced that it's cutting about $8 million in aid to Cambodia

THE ADMINISTRATION WHOSE OFFICE IS IN THE WHITE HOUSE



Metonymy

Nouns used to denote	can be used metonymically to refer to
Author (***Jane Austen*** *wrote Emma*)	**Works of Author** (*I really love* ***Jane***
Tree (***Plums*** *have beautiful*	**Fruit** (*I ate a preserved* ***plum*** *yesterday)*
Animal (*Alice Meets White* ***Rabbit***)	**Meat/Fur** (*Mary had* ***rabbit*** *for lunch)*
Place (*I haven't been to this* ***city***	**People** (*The* ***city*** *voted for Jones)*
Publication (*A daily tabloid*	**Publisher** (*The* ***newspaper*** *opposed the*

How many senses?

- difficult question…
- no hard threshold to distinguish between separate senses or closely related usages
- Kilgarriff (1997): *don't believe in word senses*
  - lexicographers: **splitters**, high number of fine-grained senses in dictionaries, good for language learning
  - computational semanticists: **lumpers**, no need of fine sense distinctions for NLP applications, group/cluster senses

WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: pattern Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) form, shape, **pattern** (a perceptual structure) *"the composition presents problems for students of musical form"; "a visual pattern must include not only objects but the spaces between them"*
- S: (n) practice, **pattern** (a customary way of operation or behavior) *"it is their practice to give annual raises"; "they changed their dietary pattern"*
- S: (n) design, **pattern**, figure (a decorative or artistic work) *"the coach had a design on the doors"*
- S: (n) convention, normal, **pattern**, rule, formula (something regarded as a normative example) *"the convention of not naming the main character"; "violence is the rule not the exception"; "his formula for impressing visitors"*
- S: (n) **pattern** (a model considered worthy of imitation) *"the American constitution has provided a pattern for many republics"*
- S: (n) blueprint, design, **pattern** (something intended as a guide for making something else) *"a blueprint for a house"; "a pattern for a skirt"*
- S: (n) traffic pattern, approach pattern, **pattern** (the path that is prescribed for an airplane that is preparing to land at an airport) *"the traffic patterns around O'Hare are very crowded"; "they stayed in the pattern until the fog lifted"*
- S: (n) radiation pattern, radiation diagram, **pattern** (graphical representation (in polar or Cartesian coordinates) of the spatial distribution of radiation from an antenna as a function of angle)

Distinguishing senses

different
truth conditions

They rarely <u>serve</u> red meat, preferring to prepare seafood

He <u>served</u> as U.S. ambassador to Norway in 1976 and 1977

He might have <u>served</u> his time, come out and led an upstanding life.

serve as NP
different
subcategorization structure

Distinguishing senses

Zeugma

- technique to discover antagonistic meanings of a word
- put two conjoined uses in a single sentence

- Which of those flights <u>**serve**</u> breakfast?
- Does Midwest Express <u>**serve**</u> Philadelphia?
- ? Does Midwest Express <u>**serve**</u> breakfast and Philadelphia?

Relations between senses

Synonymy

couch
sofa

love
enjoy

funny
comical
amusing
comic

tutor
instructor
coach

coach
bus
autobus

Relations between senses

Synonymy

- words with (nearly) identical or meanings
- substitutable in sentences without changing their truth conditions
- refer to/denote the same thing in the real world

- I wanna buy a brown leather couch / sofa

Relations between senses

Synonymy

- approximate/rough/near synonymy
- no absolute synonyms (Edmonds and Hirst, 2002)
- propositional meaning + other facets of meaning such as **style** and **connotations**

Whether the child gains the citizenship of its <u>mother</u> / *<u>mom</u>* depends upon the laws of the nation in which she is a citizen

Relations between senses

Synonymy

A relation between senses, not words

- How big/large is the plane?
- She is my big/older/elder sister

Relations between senses

??

love ⟷ hate

dark ⟷ light

arrive ⟷ leave

big ⟷ little

cold ⟷ hot

shout ⟷ whisper

above ⟷ below

tall ⟷ short

Relations between senses

Antonymy

love ⟷ hate

arrive ⟷ leave

dark ⟷ light

big ⟷ little

cold ⟷ hot

shout ⟷ whisper

above ⟷ below

tall ⟷ short

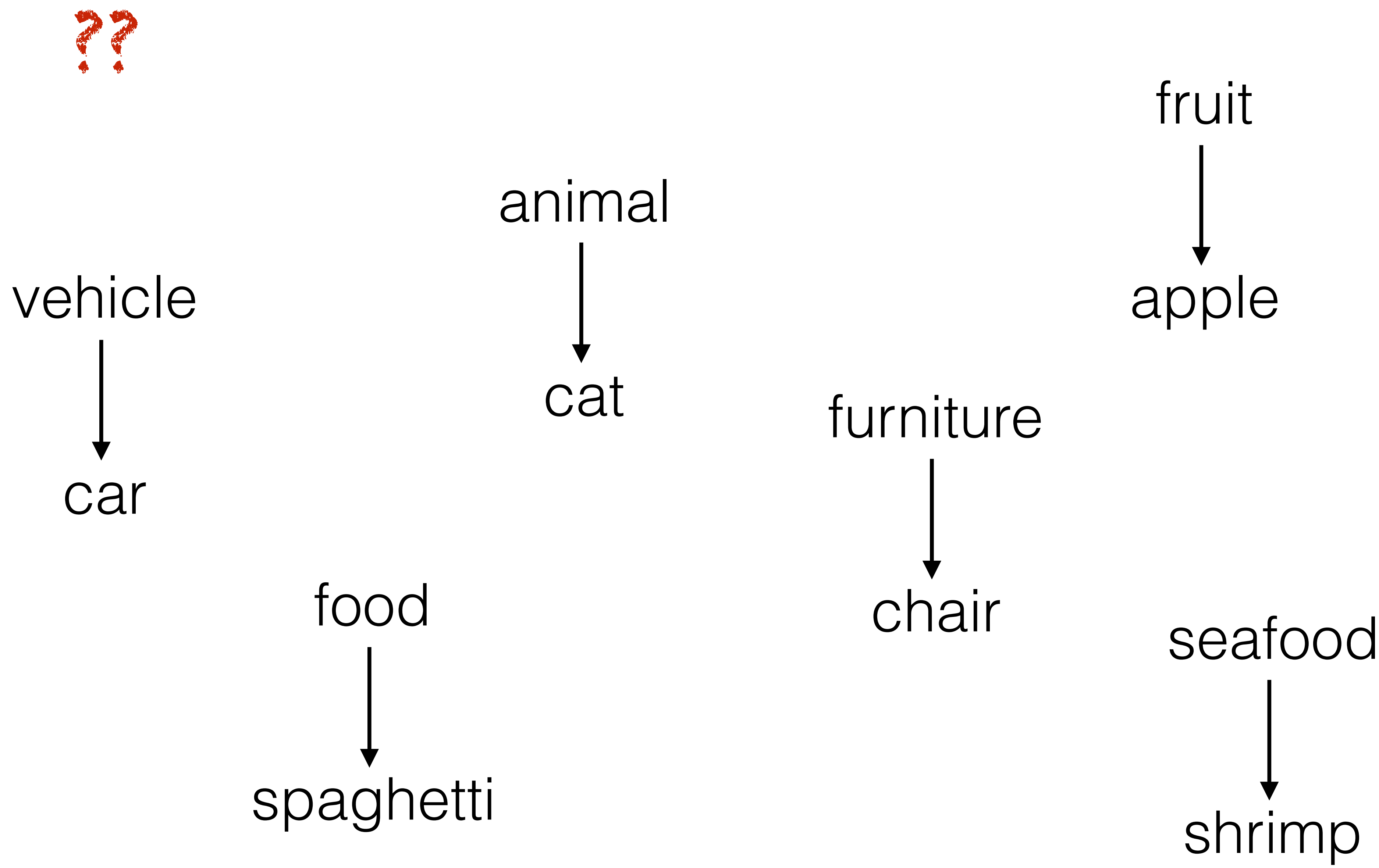
Relations between senses

Antonymy

- words with opposite meaning
- at opposite ends of a scale (e.g. length, size, temperature)
- difficult to distinguish from synonyms

- The weather was too hot/cold for me.
- I loved/hated his book.
- She is too tall/short for her age.

Relations between senses



Relations between senses

Hyponymy

fruit → apple

animal → cat

vehicle → car

furniture → chair

food → spaghetti

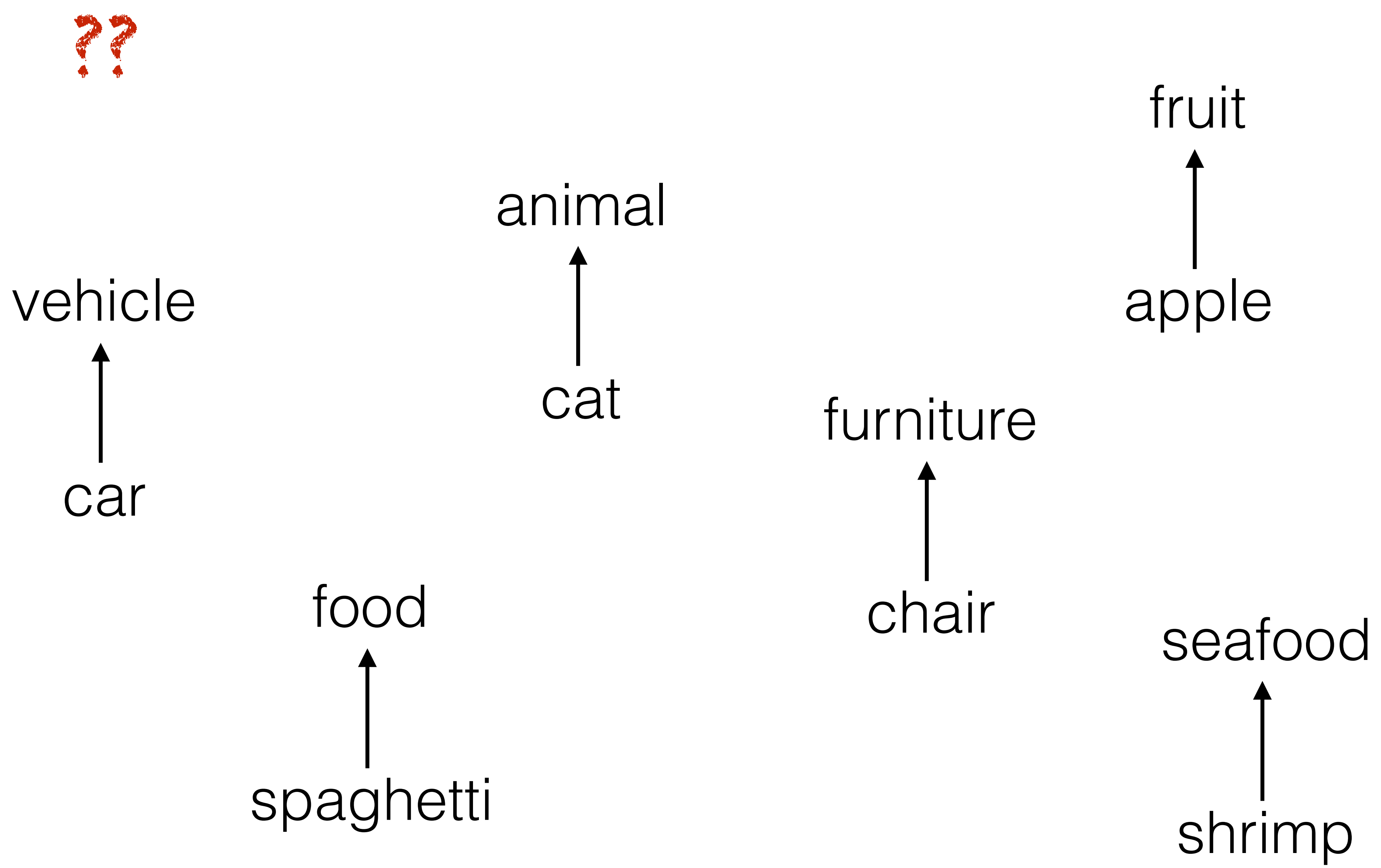
seafood → shrimp

Relations between senses

Hyponymy

- a sense A is a hyponym of a sense B if A is more **specific**/ denotes a **subclass** of B
- **IS-A** relation

Relations between senses



Relations between senses

Hypernymy

car → vehicle

cat → animal

spaghetti → food

chair → furniture

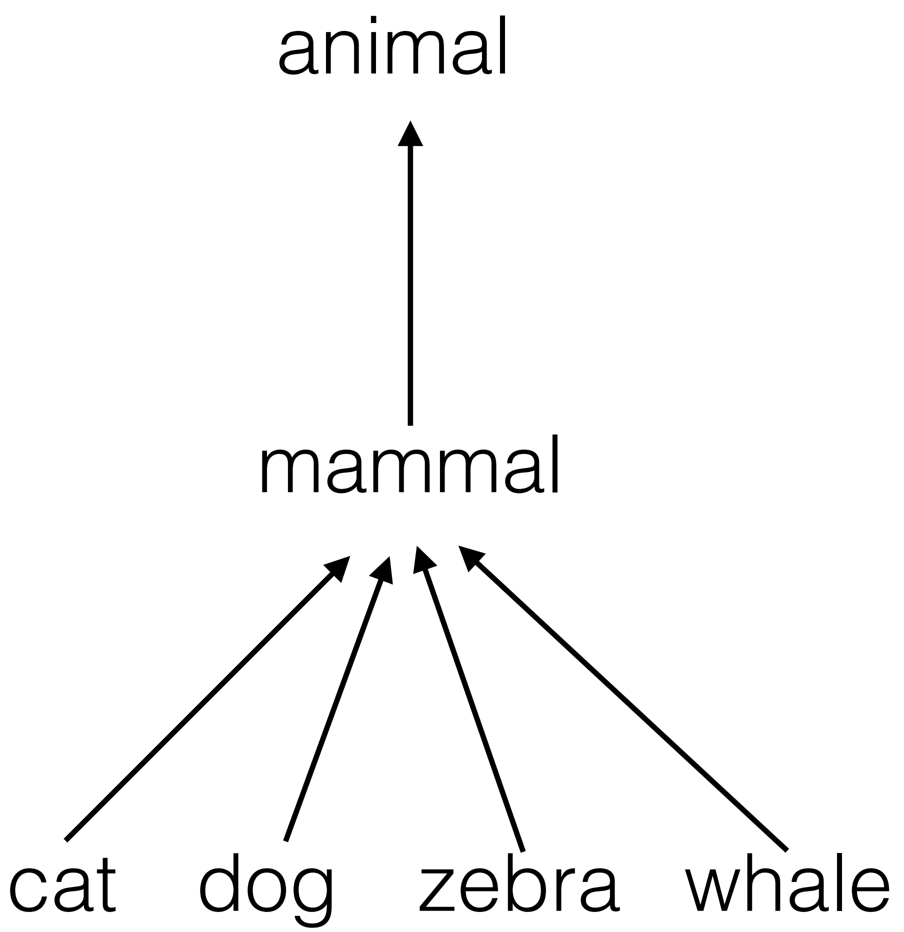
apple → fruit

shrimp → seafood

Relations between senses

Hypernymy

- the class denoted by the **superordinate** term
- hypernym includes the class denoted by the hyponym



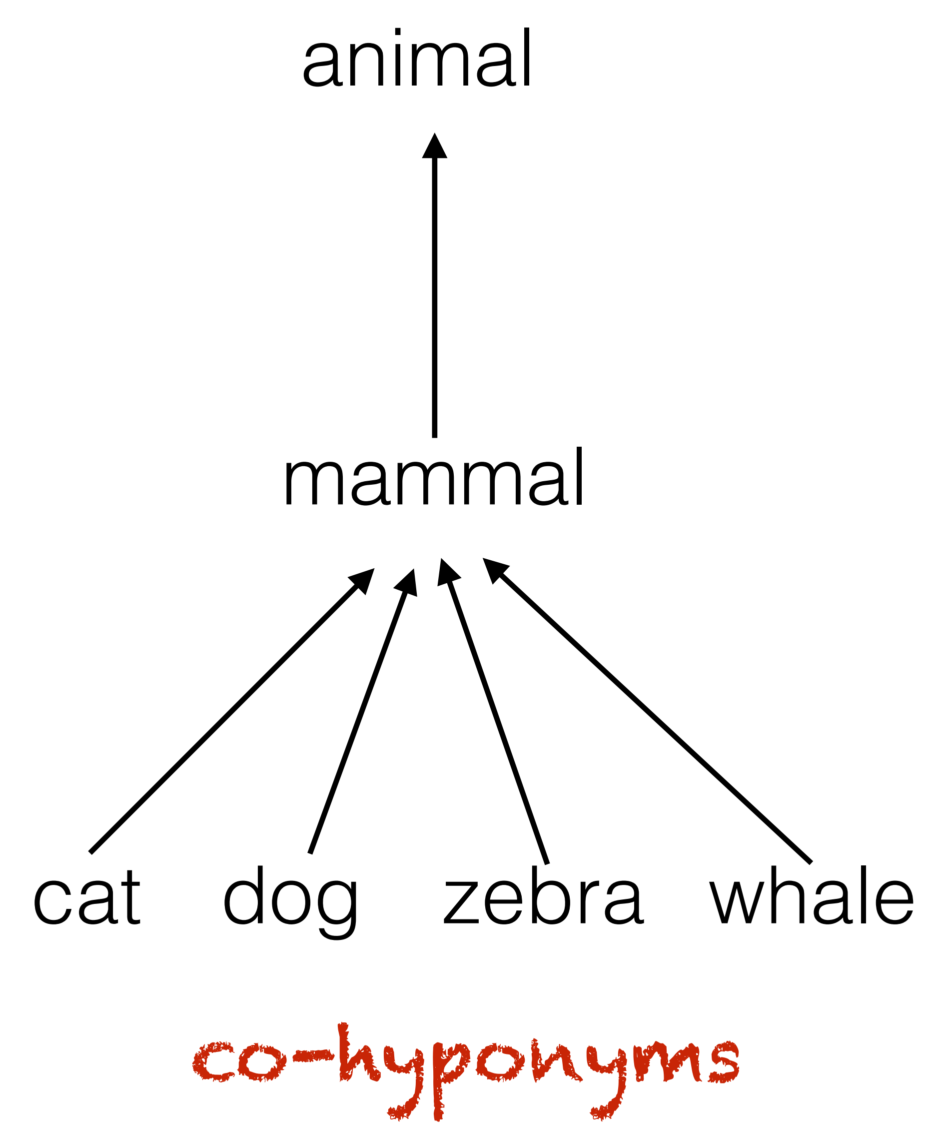
co-hyponyms

Relations between senses

Entailment

- a sense A is a hyponym of a sense B, if everything that is A is also B
- hence being an A entails being a B

$$\forall x \; A(x) \Rightarrow B(x)$$



Relations between senses

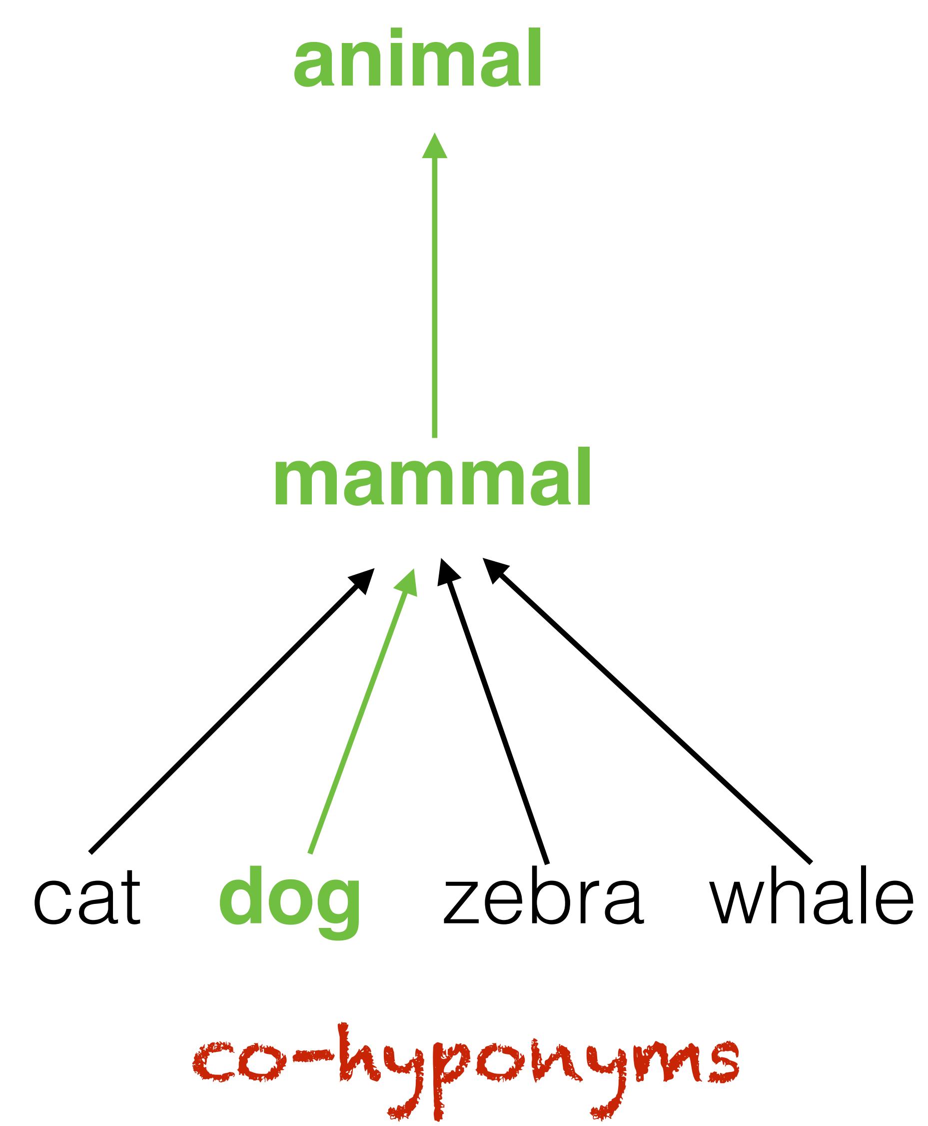
Entailment

- a sense A is a hyponym of a sense B, if everything that is A is also B
- hence being an A entails being a B

$$\forall x \; A(x) \Rightarrow B(x)$$

Transitivity

- if A is a hyponym of B and B is a hyponym of C, then A is a hyponym of C



Relations between senses

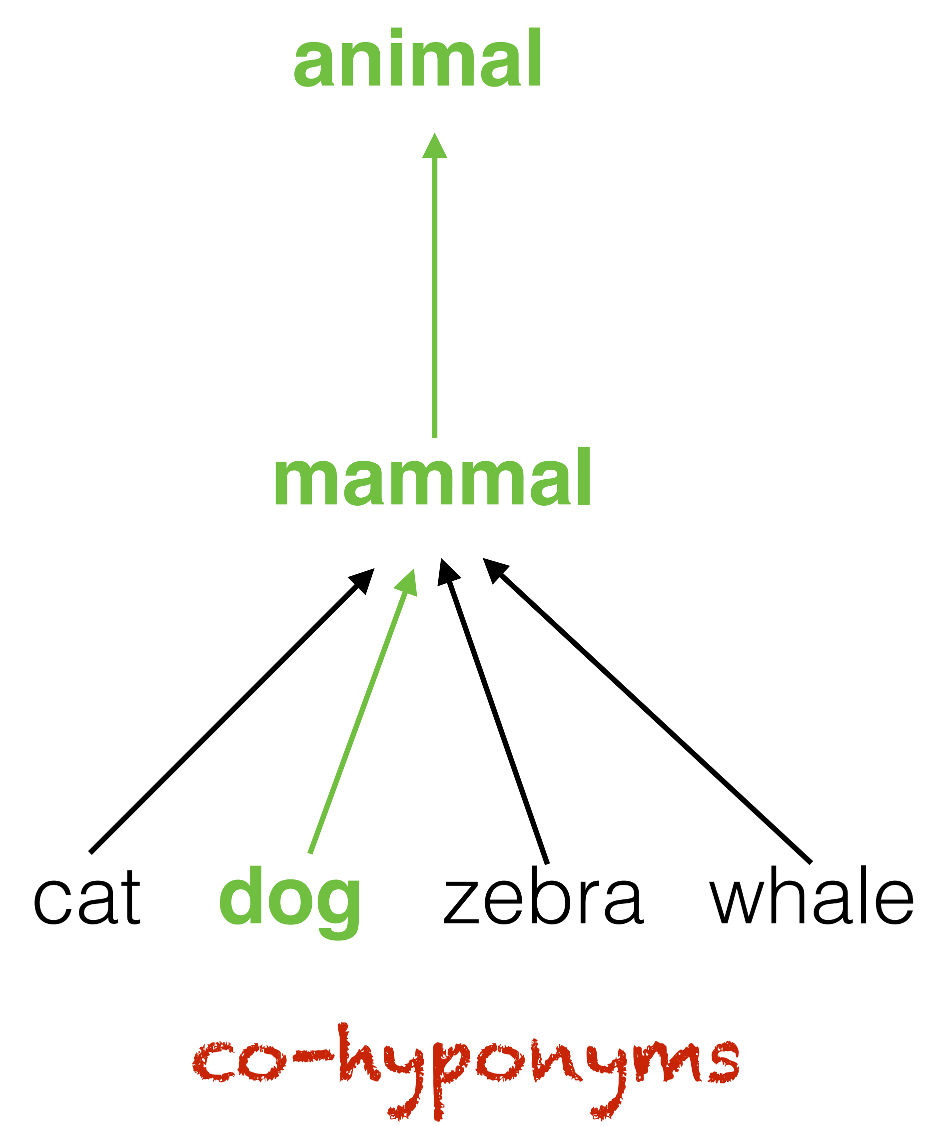
Entailment

- a sense A is a hyponym of a sense B, if everything that is A is also B
- hence being an A entails being a B

$$\forall x \; A(x) \Rightarrow B(x)$$

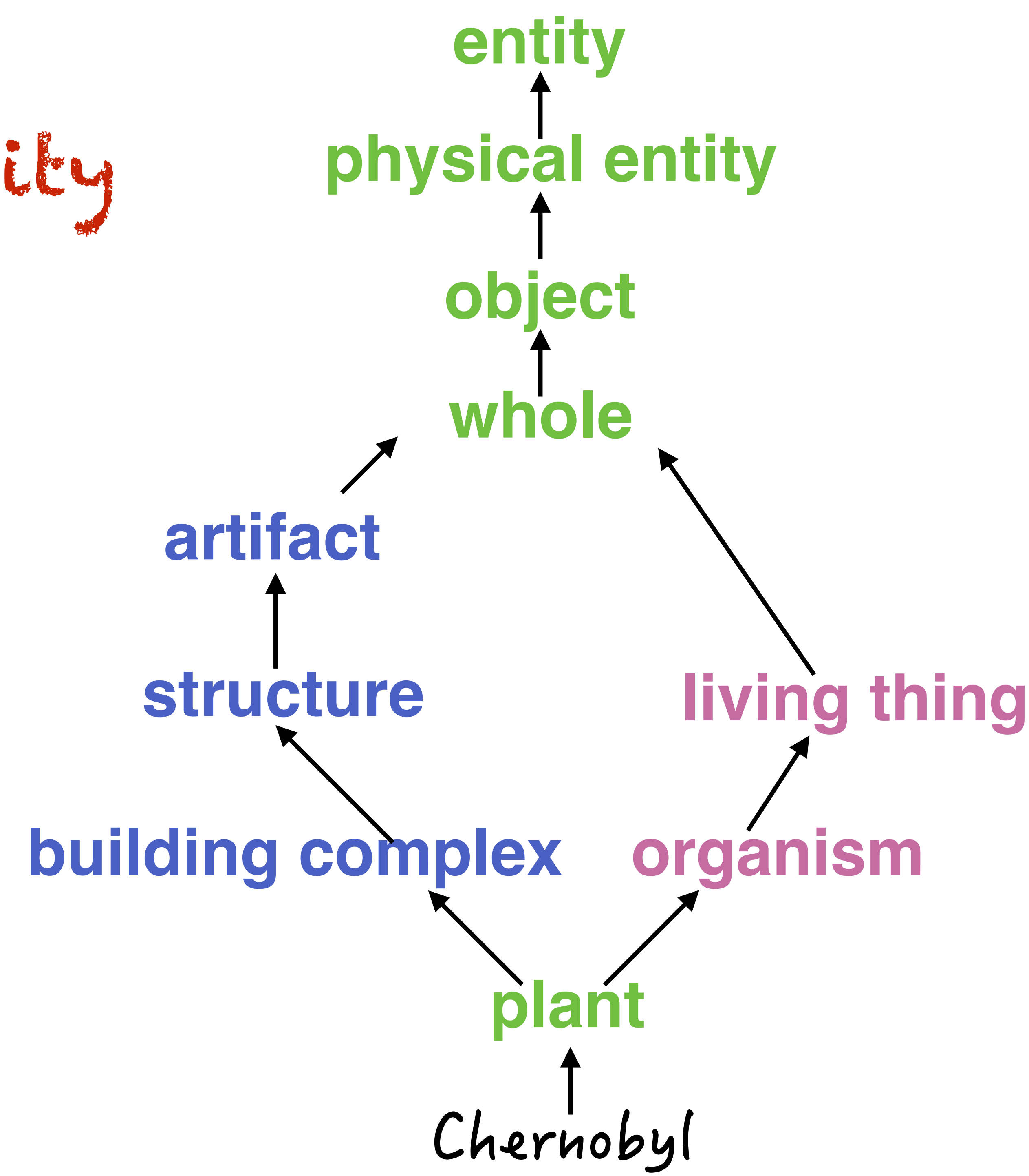
Transitivity

- if A is a hyponym of B and B is a hyponym of C, then A is a hyponym of C



Relations between senses

Transitivity



Relations between senses

??

chair → leg

cat → whisker

knife → blade

flower → petal

tv → screen

car → wheel

chicken → wing

book → cover

Relations between senses

Meronymy (part-whole relation)

chair → leg

cat → whisker

knife → blade

flower → petal

tv → screen

car → wheel

book → cover

chicken → wing

Relations between senses

Holonymy

chair ← leg

cat ← whisker

knife ← blade

flower ← petal

tv ← screen

car ← wheel

book ← cover

chicken ← wing

WordNet

- a database of lexical relations (Fellbaum, 1998)
  - 117,798 nouns
  - 11,529 verbs
  - 22,479 adjectives
  - 4,481 adverbs

WordNet

coach

glosses

synsets

Noun

- S: (n) **coach**, manager, handler ((sports) someone in charge of training an athlete or a team)
- S: (n) **coach**, private instructor, tutor (a person who gives private instruction (as in singing, acting, etc.))
- S: (n) passenger car, **coach**, carriage (a railcar where passengers ride)
- S: (n) **coach**, four-in-hand, coach-and-four (a carriage pulled by four horses with one driver)
- S: (n) bus, autobus, **coach**, charabanc, double-decker, jitney, motorbus, motorcoach, omnibus, passenger vehicle (a vehicle carrying many passengers; used for public transport) *"he always rode the bus to work"*

examples

Verb

- S: (v) **coach**, train (teach and supervise (someone); act as a trainer or coach (to), as in sports) *"He is training our Olympic team"; "She is coaching the crew"*
- S: (v) **coach** (drive a coach)

WordNet

Synsets (synonym sets) - concepts

coach

noun

1. coach^1, manager^2, handler^3
2. coach^2, private instructor^1, tutor^1
3. passenger car^1, coach^3, carriage^1
4. coach^4, four-in-hand^2, coach-and-four
5. bus^1, autobus^1, coach^5, charabanc^1, double-decker^1, jitney^1, motorbus^1, motorcoach^1, omnibus^2, passenger vehicle^1

verb

1. coach^2, train^7
2. coach^2

Hypernymy chain for coach - Sense 1

coach[1], **manager**[2], **handler**[3] ((sports) someone in charge of training an athlete or a team)

=> **trainer** (one who trains other persons or animals)

=> **leader** (a person who rules or guides or inspires others)

=> **person**, **individual**, **someone**, **somebody**, **mortal**, **soul** (a human being) "there was too much for one person to do"

=> **organism**, **being** (a living thing that has (or can develop) the ability to act or function independently)

=> **living thing**, **animate thing** (a living (or once living)entity)

=> **whole**, **unit** (an assemblage of parts that is regarded as a single entity) "how big is that part compared to the whole?"; "the team is a unit"

=> **object**, **physical object** (a tangible and visible entity; an entity that can cast a shadow) "it was full of rackets, balls and other objects"

=> **physical entity** (an entity that has physical existence)

=> **entity** (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Hypernymy chain for coach - Sense 5

bus[1], **autobus**[1], **coach**[5], **charabanc**[1], **double-decker**[1], **jitney**[1], **motorbus**[1], **motorcoach**[1], **omnibus**[2], **passenger vehicle**[1]

=> **public transport** (conveyance for passengers or mail or freight)

- => **conveyance, transport** (something that serves as a means of transportation)
  - => **instrumentality, instrumentation** (an artifact (or system of artifacts) that is instrumental in accomplishing some end)
    - => **artifact, artefact** (a man-made object taken as a whole)
      - => **whole, unit** (an assemblage of parts that is regarded as a single entity) “how big is that part compared to the whole?”; “the team is a unit”
        - => **object, physical object** (a tangible and visible entity; an entity that can cast a shadow) “it was full of rackets, balles and other objects”
          - => **physical entity** (an entity that has physical existence)
            - => **entity** (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

All relations in WordNet

Nouns

Hypernym/Hyponym: furniture -> chair -> armchair

(Member/Part/Substance) Meronym/Holonym:

chair -> back, seat -> leg

Instance Hypernym/Hyponym: president -> Barack Obama

Adjectives

Direct Antonyms: wet - dry

Indirect antonyms: parched/arid/dessicated/bone-dry — soggy, waterlogged

Relational adjectives (pertainyms): criminal -> crime

Verbs

Troponyms: communicate - talk - whisper

Antonyms: whisper - shout

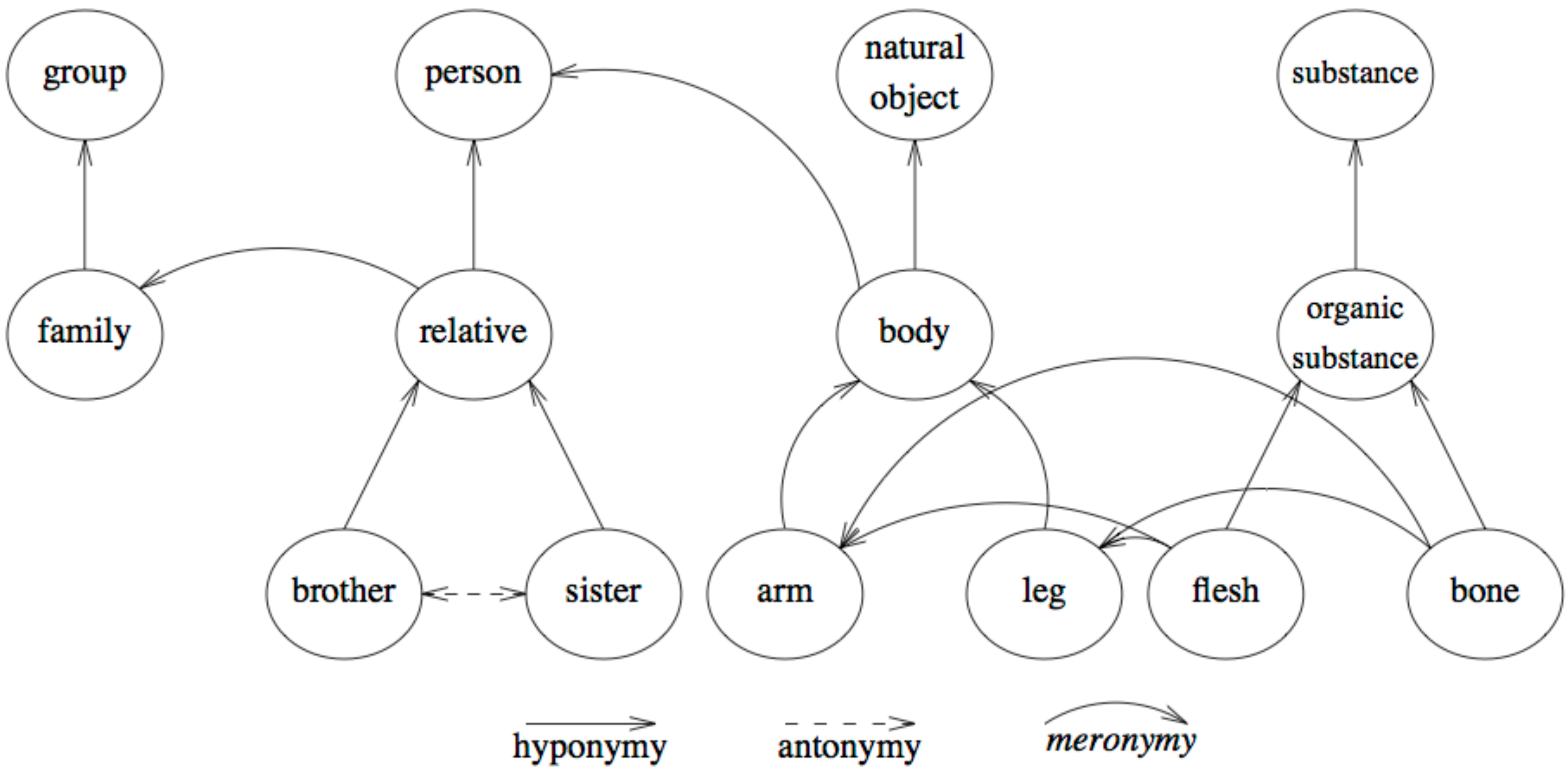
Speed: move-jog-run

Intensity of emotion: like - love - idolize

Entailment: buy-pay, succeed-try, show-see

Derivationally related: whisper-whisperer, whispering

WordNet graph



Part 2
Word Sense Disambiguation

If one examines the words in a book, one at a time through an opaque mask with a hole in it one word wide, then it is obviously impossible to determine, one at a time, the meaning of words. " Fast" may mean "rapid"; or it may mean "motionless"; and there is no way of telling which.

But, if one lengthens the slit in the opaque mask, until one can see not only the central word in question but also say N words on either side, then, if N is large enough one can unambiguously decide the meaning...

Warren Weaver's "Translation" memorandum (1949)

Word Sense Disambiguation

The task of selecting the correct sense for a word in context

The Super Bowl will be the first time that Doug Pederson will match wits with and try to outscheme Bill Belichick as a head coach.

[TRAINER, TUTOR, INSTRUCTOR] ?

[BUS, AUTOBUS, CARRIAGE] ?

Word Sense Disambiguation

Machine Translation

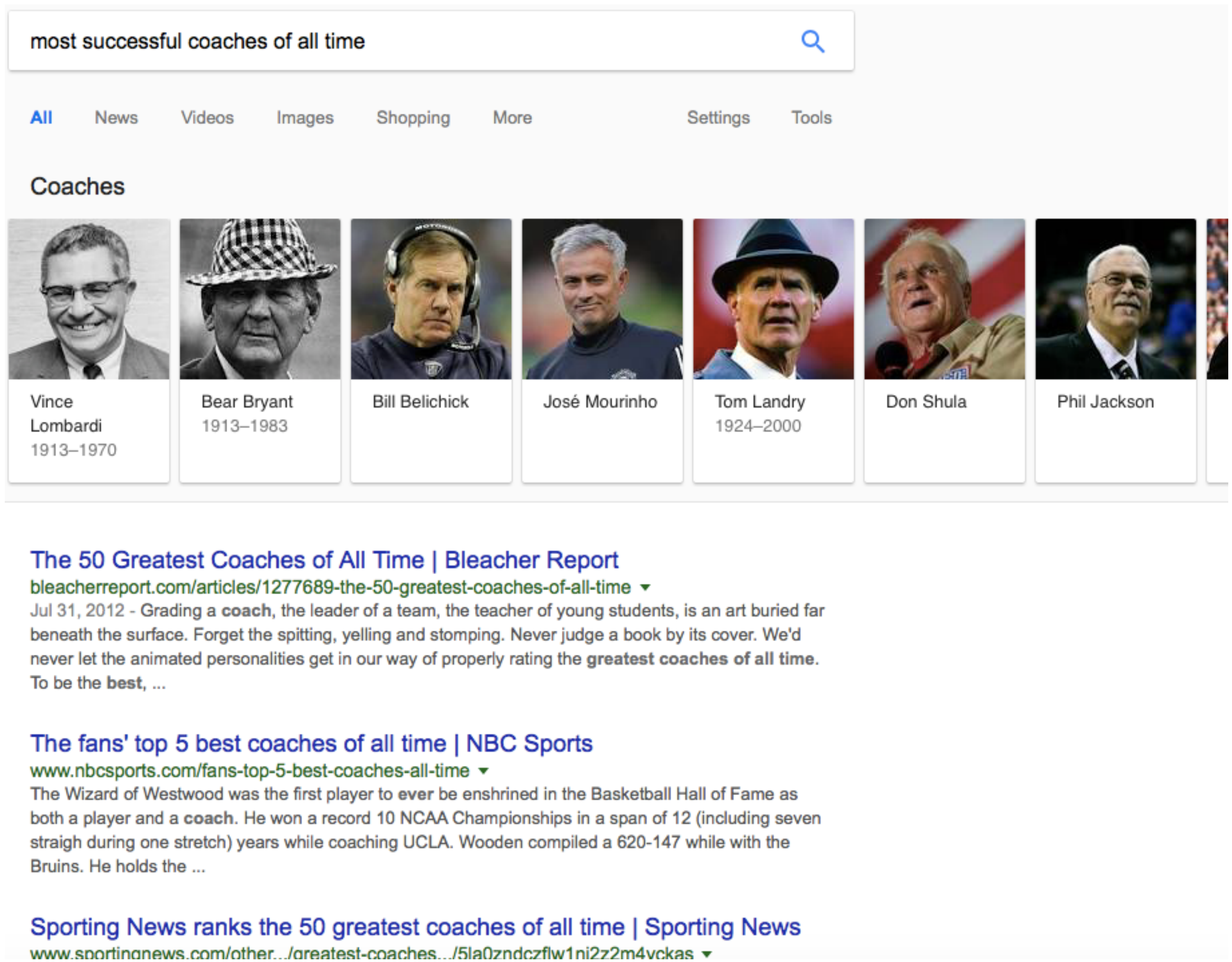
as a head **coach** -> en tant qu'**entraîneur** principal (autobus, wagon)
-> **entrenador** en jefe (autobus, coche, vagón)

Question Answering

Who is the Eagles **coach** -> Doug Pederson

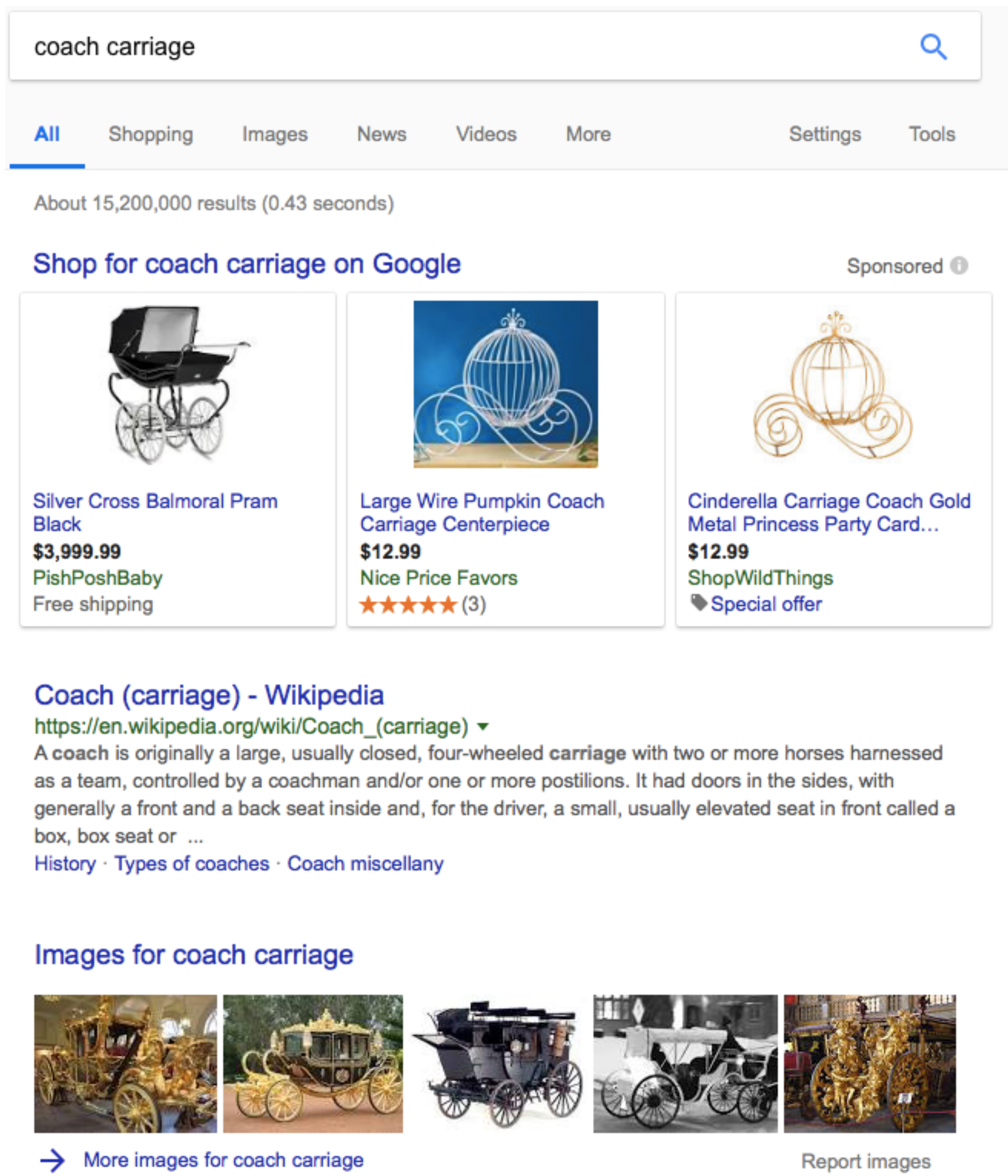
Word Sense Disambiguation

Information Retrieval



Word Sense Disambiguation

Information Retrieval



Word Sense Disambiguation

The Super Bowl will be the first time that Doug Pederson will match wits with and try to outscheme Bill Belichick as a head <u>coach</u>.

WSD algorithm

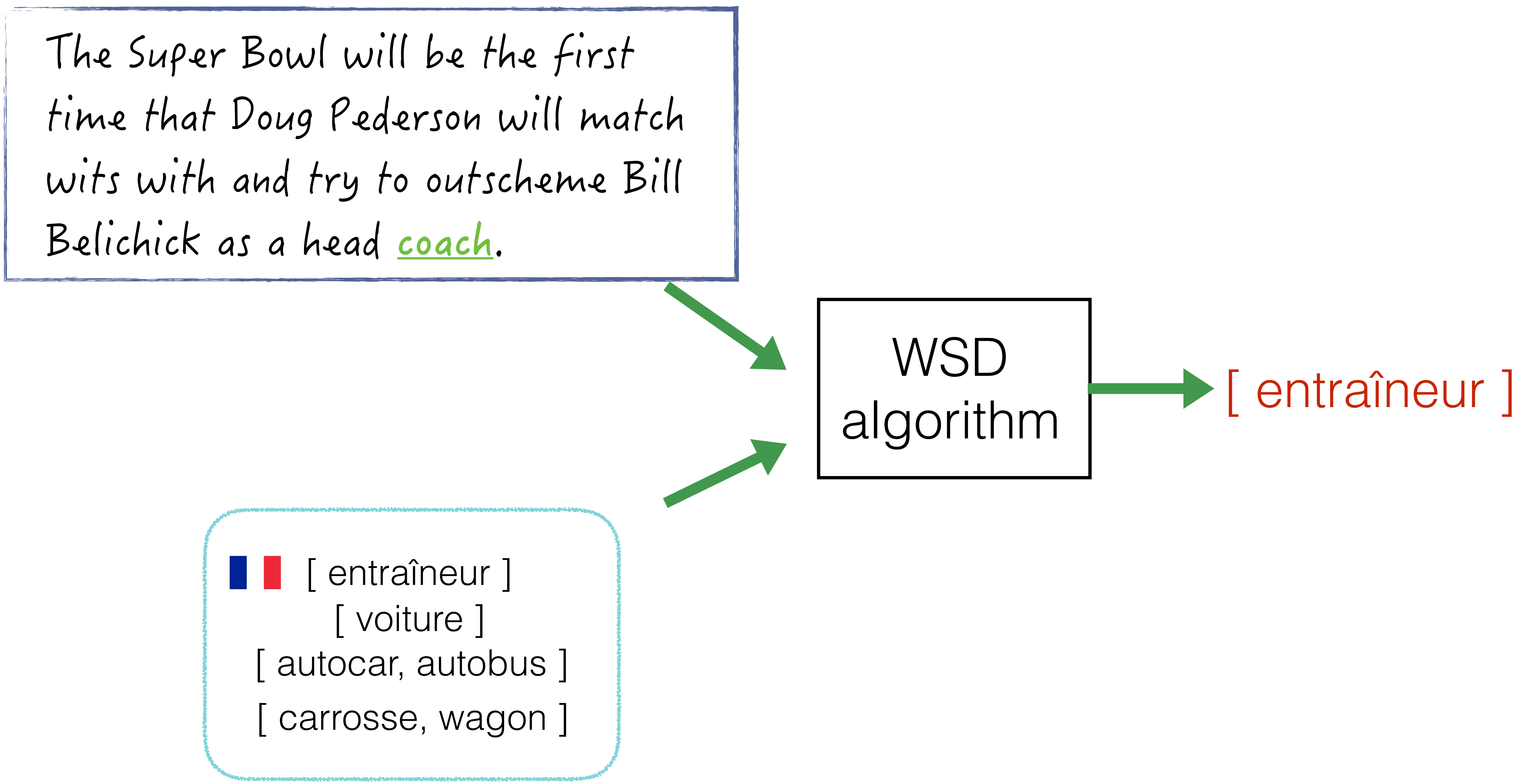
[COACH, MANAGER, HANDLER]

[COACH, MANAGER, HANDLER]

[PASSENGER CAR, COACH, CARRIAGE]

[BUS, AUTOBUS, COACH, CHARABANC, DOUBLE-DECKER,JITNEY, MOTORBUS, MOTORCOACH, OMNIBUS, PASSENGER VEHICLE]

Word Sense Disambiguation



Supervised WSD

- **training data hand-labeled** with word senses (e.g. SemCor corpus, data from SemEval tasks)
- extraction of features predictive of word senses
- classifier trained to assign the correct sense given these features
- output of training => a classifier system capable of assigning sense labels to **unlabeled words in context**

2 flavors: Lexical sample / All Words

Supervised WSD - Lexical Sample

This devolution will take place amid bitter conflict over local electoral and citizenship laws , and an explosive <head>**argument**</head> over the very legality of the incorporation of Estonia into the Soviet Union in 1940 .

After an <head>**argument**</head> about politics and poetics on the verandah of the Norfolk Hotel , I would often go with Jenni to the movies .

The latter camp took heart yesterday from Mr Rocard 's insistence that any decision must be based on a consensus , but realise that they have a tough fight ahead persuading others of the logic of their <head>**arguments**</head> .

Without incentives , goes the <head>**argument**</head> , the drift of talented young players towards more successful counties , with open doors , will accelerate .

Supervised WSD - Lexical Sample

This devolution will take place amid bitter conflict over local electoral and citizenship laws , a socially explosive <head>**argument**</head> over the alleged illegality of the incorporation of Estonia into the Soviet Union in 1940 .

[controversy contention contestation tilt argument arguing]

After an <head>**argument**</head> about politics and poetics on the verandah of the Norfolk Hotel , I would often go with Jenni to the movies .

[argument debate]

The latter camp took heart yesterday from Mr Rocard 's insistence that any decision must be based on a consensus , but realise that they have a tough fight ahead persuading others of the logic of their <head>**arguments**</head> .

[argument, statement]

Without incentives , goes the <head>**argument**</head> , the drift of talented young players towards more successful counties , with open doors , will accelerate .

[argument, statement]

Supervised WSD - All Words

- all open-class words in a sentence labeled with a word sense
- **SemCor** corpus
  - a subset of the Brown Corpus
  - over 234,000 words manually tagged with WordNet senses
- **SensEval-3** English All-words test data
  - 2,081 tagged content word tokens from the WSJ and Brown corpora

Supervised WSD - All Words

SemCor

<wf cmd=ignore pos=PRP$>**His**</wf>
<wf cmd=done pos=NN lemma=father **wnsn=1** lexsn=1:18:00::>**father**</wf>
<wf cmd=done rdf=person pos=NNP lemma=person **wnsn=1** lexsn=1:03:00:: pn=person>**Soeren**</wf>
<wf cmd=done pos=VB lemma=be **wnsn=2** lexsn=2:42:06::>**was**</wf>
<wf cmd=ignore pos=DT>**the**</wf>
<wf cmd=done pos=NN lemma=village **wnsn=1** lexsn=1:14:00::>**village**</wf>
<wf cmd=done pos=NN lemma=apothecary **wnsn=1** lexsn=1:18:00::>**apothecary**</wf>
<wf cmd=ignore pos=WP$>**whose**</wf>
<wf cmd=done pos=JJ lemma=slender **wnsn=4** lexsn=5:00:00:small:00>**slender**</wf>
<wf cmd=done pos=NN lemma=income **wnsn=1** lexsn=1:21:00::>**income**</wf>
<wf cmd=done pos=VB lemma=make **wnsn=2** lexsn=2:30:00::>**made**</wf>
<wf cmd=done pos=VB ot=notag>**it**</wf>
<wf cmd=done pos=JJ lemma=difficult **wnsn=1** lexsn=3:00:00::>**difficult**</wf>
<wf cmd=ignore pos=TO>**to**</wf>
<wf cmd=done pos=VB lemma=feed **wnsn=2** lexsn=2:34:01::>**feed**</wf>
<wf cmd=ignore pos=PRP$>**his**</wf>
<wf cmd=done pos=NN lemma=family **wnsn=1** lexsn=1:14:02::>**family**</wf>
<punc>,</punc>
<wf cmd=done pos=RB lemma=let_alone **wnsn=1** lexsn=4:02:00::>**let_alone**</wf>
<wf cmd=done pos=VB lemma=educate **wnsn=1** lexsn=2:41:00::>**educate**</wf>
<wf cmd=ignore pos=PRP>**them**</wf>
…

Supervised WSD - All Words

SemCor

<wf cmd=ignore pos=PRP$>**His**</wf>

male parent, begetter

<wf cmd=done pos=NN lemma=father **wnsn=1** lexsn=1:18:00::>**father**</wf>

<wf cmd=done rdf=person pos=NNP lemma=person **wnsn=1** lexsn=1:03:00:: pn=person>**Soeren**</wf>

<wf cmd=done pos=VB lemma=be **wnsn=2** lexsn=2:42:06::>**was**</wf>

<wf cmd=ignore pos=DT>**the**</wf>

small town, settlement

<wf cmd=done pos=NN lemma=village **wnsn=1** lexsn=1:14:00::>**village**</wf>

<wf cmd=done pos=NN lemma=apothecary **wnsn=1** lexsn=1:18:00::>**apothecary**</wf>

pharmacist, druggist, chemist

<wf cmd=ignore pos=WP$>**whose**</wf>

<wf cmd=done pos=JJ lemma=slender **wnsn=4** lexsn=5:00:00:small:00>**slender**</wf> slim

<wf cmd=done pos=NN lemma=income **wnsn=1** lexsn=1:21:00::>**income**</wf>

<wf cmd=done pos=VB lemma=make **wnsn=2** lexsn=2:30:00::>**made**</wf> get

<wf cmd=done pos=VB ot=notag>**it**</wf>

<wf cmd=done pos=JJ lemma=difficult **wnsn=1** lexsn=3:00:00::>**difficult**</wf> hard

<wf cmd=ignore pos=TO>**to**</wf>

<wf cmd=done pos=VB lemma=feed **wnsn=2** lexsn=2:34:01::>**feed**</wf>

<wf cmd=ignore pos=PRP$>**his**</wf>

household, house, home, menage

<wf cmd=done pos=NN lemma=family **wnsn=1** lexsn=1:14:02::>**family**</wf>

<punc>,</punc>

<wf cmd=done pos=RB lemma=let_alone **wnsn=1** lexsn=4:02:00::>**let_alone**</wf>not to mention

<wf cmd=done pos=VB lemma=educate **wnsn=1** lexsn=2:41:00::>**educate**</wf>

<wf cmd=ignore pos=PRP>**them**</wf>

…

Supervised WSD

- **pre-processing**: lemmatization, part-of-speech tagging, syntactic parsing
- extract **context features** relevant to the target word
- construct a **feature vector** which encodes this linguistic information, to use as input to a machine learning algorithm

Features

Bag of words feature vector

- unordered set of words
- vector of binary features indicating whether a vocabulary word *w* occurs or not in the context

He is the winningest coach in American University history

Features

Collocational features

- encode information about **specific positions** located to the left and right of the target word (e.g. 1 word to the right/left)

Typical features

- the word itself, its root form, its POS

Collocational feature vector

He is the winningest coach in American University history

$$[w_{i-2}, POS_{i-2}, w_{i-1}, POS_{i-1}, w_{i+1}, POS_{i+1}, w_{i+2}, POS_{i+2}]$$

[the, DET, winning, ADJ, in, IN, American, ADJ]

WSD Evaluation

Extrinsic / task-based / end-to-end evaluation

- does WSD improve performance in some application (e.g. question answering, MT) ?

Intrinsic evaluation

- measure sense prediction accuracy, Precision, Recall
- SensEval/SemEval datasets, test portion of SemCor

Baseline

- the most frequent sense (MFS): powerful heuristic because of the skewed distribution of word senses

Knowledge-based WSD

- sense-labeled data needed for supervised WSD is expensive and limited
- indirect supervision from dictionaries, thesauruses, and other ressources
- knowledge-based or **weakly supervised WSD**

Dictionary-based WSD

The Lesk Algorithm (1986)

- compare the sense glosses (definitions) of a word with the glosses of each of the context words
- choose the sense whose gloss shares the most words with glosses of the context words

How to Tell a Pine Cone from an Ice Cream Cone?

pine 1 kinds of **evergreen tree** with needle-shaped leaves
2 waste away through sorrow and illness

cone 1 solid body which narrows to a point
2 something of this shape whether solid or hollow
3 fruit of certain **evergreen trees**

Dictionary-based WSD

Simplified Lesk (Kilgarriff and Rosenzweig, 2000)

- compare the glosses of a word with the context words
- choose the sense whose gloss shares the most words with the context

```
function SIMPLIFIED LESK(word, sentence) returns best sense of word

  best-sense ← most frequent sense for word
  max-overlap ← 0
  context ← set of words in sentence
  for each sense in senses of word do
    signature ← set of words in the gloss and examples of sense
    overlap ← COMPUTEOVERLAP(signature, context)
    if overlap > max-overlap then
         max-overlap ← overlap
         best-sense ← sense
  end
  return(best-sense)
```

Dictionary-based WSD

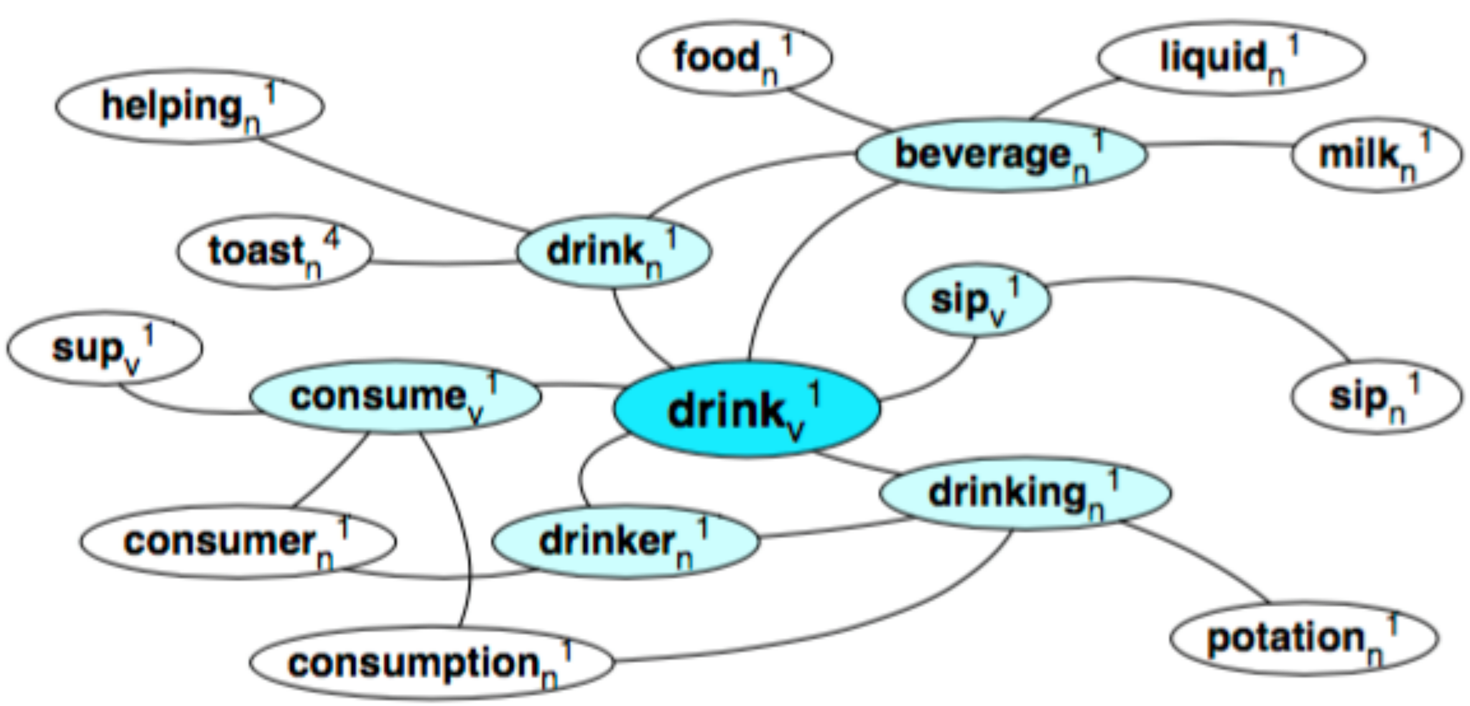
Corpus Lesk (Vasilescu et al., 2004)

- glosses are short, not much overlap with the context/other entries
- **expand sense definitions** to include related words
- adds in the signature for one sense, all the other words tagged with the same sense in a corpus
- apply a weight to each overlapping word (e.g. using TF-IDF; how many other glosses the word occurs in; discount function words)

Graph-based WSD

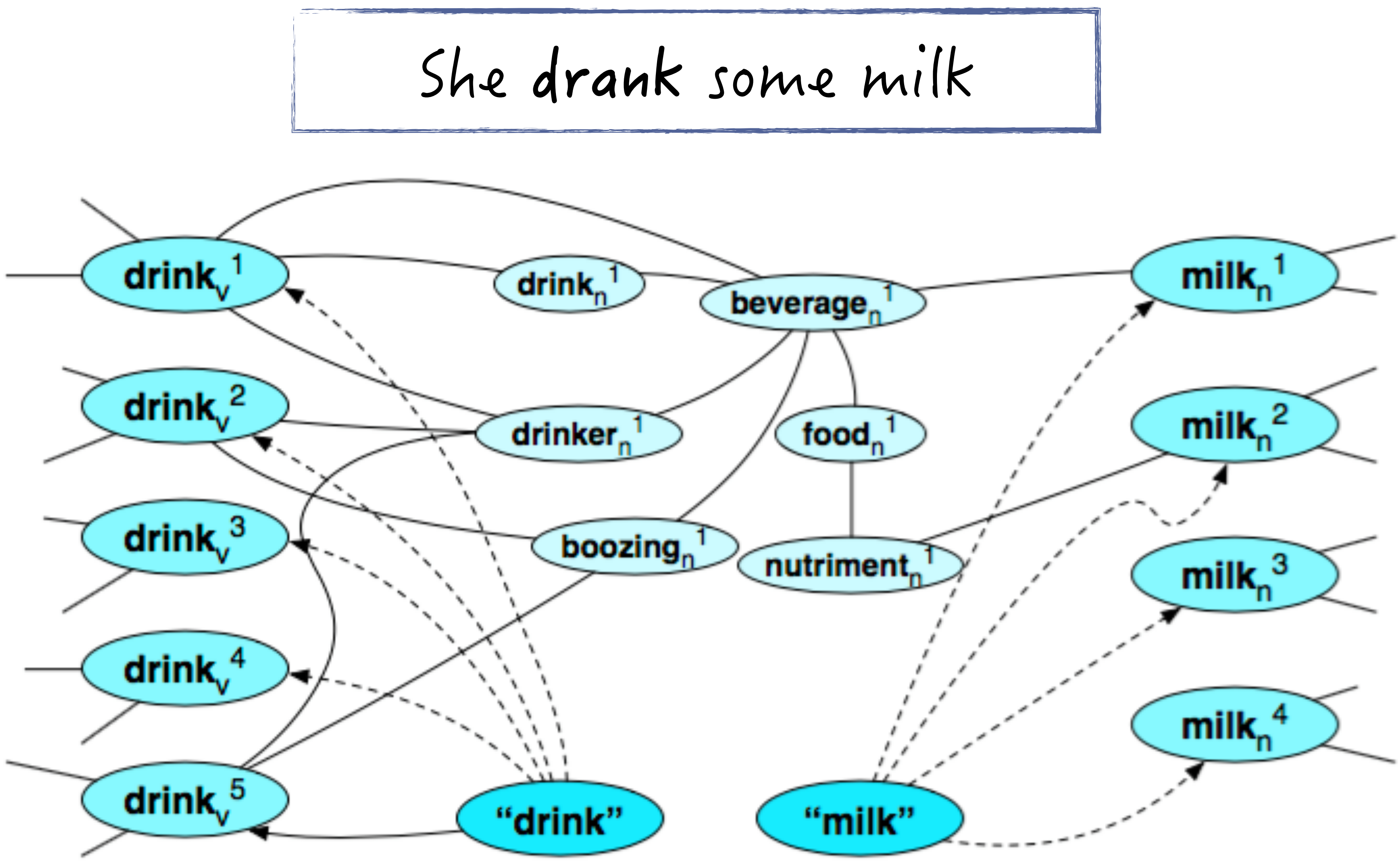
WordNet can be construed as a **graph**

- nodes => senses, edges => relations



Navigli and Lapata (2010) integrate standard WordNet relations + relations between senses and unambiguous words in glosses

Graph-based WSD



- the target word and the words in its context are inserted in the graph
- the correct sense is the most central in the graph (degree, Personalized Page Rank)

Semi-Supervised WSD

The Yarowsky algorithm (1995) uses **bootstrapping**

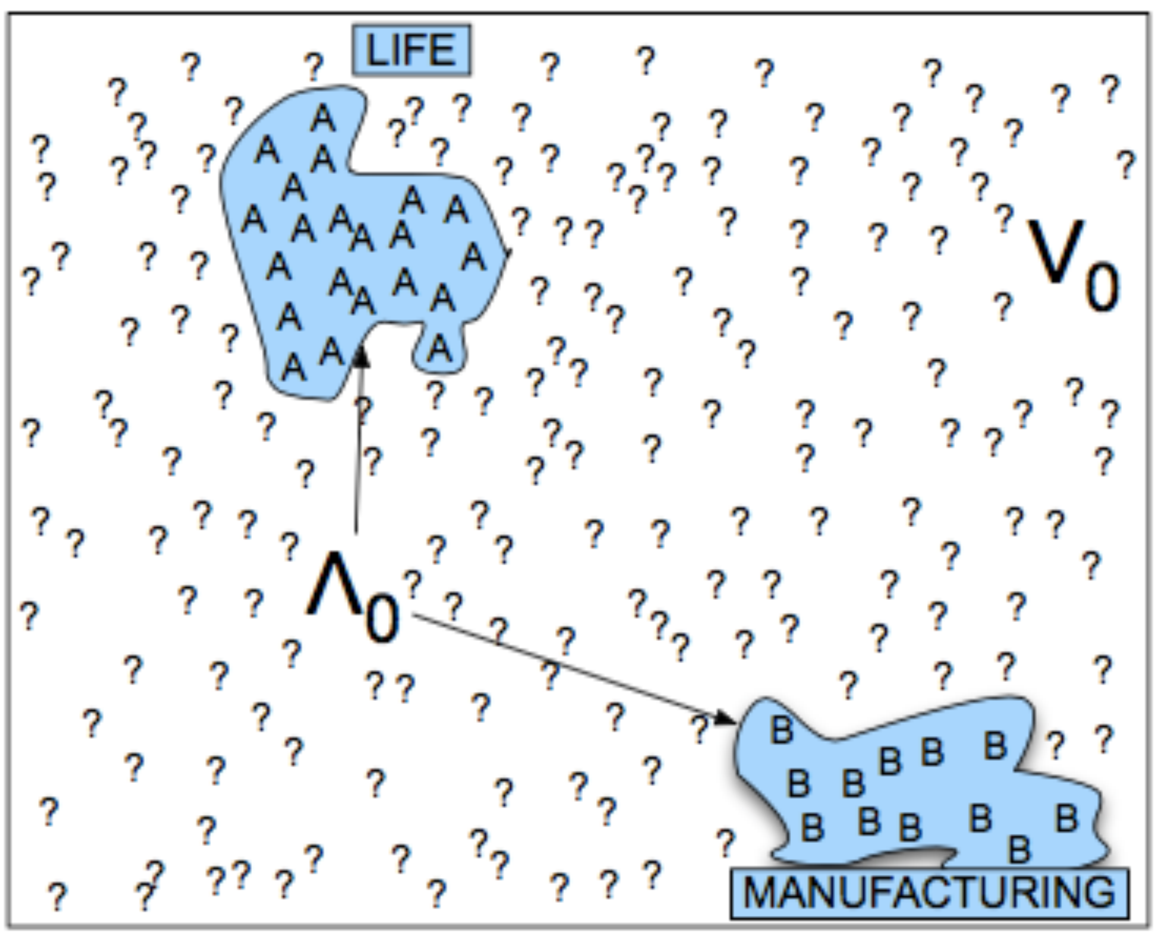
- small seedset of sense labeled instances & large unlabeled corpus
- classifier trained on seedset, used to label the unlabeled corpus
- select examples with high confidence, add them to the training set
- train new classifier, apply to unlabeled set, extract new training set, etc

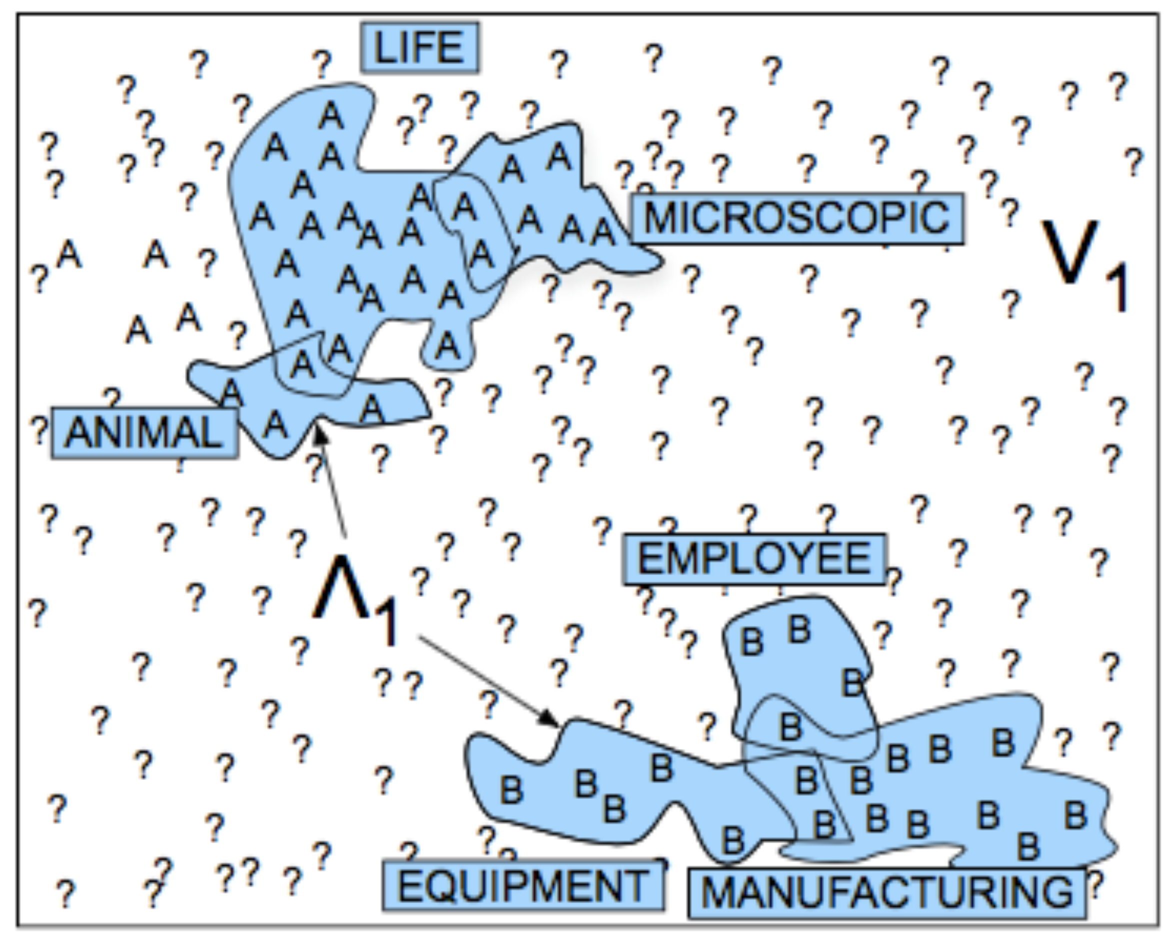
With each iteration the training corpus **grows**, the untagged corpus **shrinks**

Process repeated until no new examples are above threshold

Semi-Supervised WSD

plant





Semi-Supervised WSD

Initial seeds

- small set of hand-labeled examples or discovered using a heuristic

One-sense per collocation

- nearby words provide strong and consistent clues to the sense of a target word, conditional on relative distance, order and syntactic relationship

One-sense per discourse

- the sense of a target word is highly consistent within any given document

Semi-Supervised WSD

One-sense per collocation

Initial decision list for *plant* (abbreviated)		
LogL	Collocation	Sense
8.10	***plant*** **life**	⇒ A
7.58	**manufacturing** ***plant***	⇒ B
7.39	life (within ±2-10 words)	⇒ A
7.20	manufacturing (in ±2-10 words)	⇒ B
6.27	animal (within ±2-10 words)	⇒ A
4.70	equipment (within ±2-10 words)	⇒ B
4.39	employee (within ±2-10 words)	⇒ B
4.30	assembly *plant*	⇒ B
4.10	*plant* closure	⇒ B
3.52	*plant* species	⇒ A
3.48	automate (within ±2-10 words)	⇒ B
3.45	microscopic *plant*	⇒ A
	...	

- decision-list approach
- compute word-sense probability distributions for all collocations
- order them by log-likelihood ratio

$$Log\left(\frac{\Pr(Sense_A \mid Collocation_i)}{\Pr(Sense_B \mid Collocation_i)}\right)$$

Semi-Supervised WSD

One sense per discourse (Gale et al., 1992)

- a particular word appearing multiple times in a text or discourse often appears with the same sense
- this heuristic holds better for coarse-grained senses (homonymy)
- it can improve many WSD methods

Part 3
Word Sense Induction

Automatic Sense Induction

- large sense labeled corpora are scarce
- automatic Word Sense Induction (WSI) needs no human-defined word senses
- set of senses for each word created automatically from its instances in the training set

Distributional hypothesis of meaning

- You shall know a word by the company it keeps (Firth, 1957)
- Words that occur in same contexts have similar meanings (Harris, 1954)
- Charles and Miller (1991), Wittgenstein (1953), …

Distributional hypothesis of meaning

- You shall know a word by the company it keeps (Firth, 1957)
- Words that occur in same contexts have similar meanings (Harris, 1954)
- Charles and Miller (1991), Wittgenstein (1953), …

We found a little, hairy wampimuk sleeping behind the tree

(Baroni, 2014)

A bottle of tezgüno is on the table.
Everyone likes tezgüno.
Tezgüno makes you drunk.
We make tezgüno out of corn.

(Pantel and Lin, 2002)

Distributional semantic models

- Represent words through vectors recording their co-occurrence counts with context elements in a corpus
- Re-weight counts in the matrix to give higher weight to more informative co-occurrences (e.g. Positive Pointwise Mutual Information, TF-IDF)
- Apply dimensionality reduction techniques to the matrix (e.g. Singular Value Decomposition, Non-negative matrix factorization)
- Measure geometric distance of word vectors in “distributional space” as proxy to semantic similarity

Distributional semantic models

```
he curtains open and the moon shining in on the barely
ars and the cold , close moon " . And neither of the w
rough the night with the moon shining so brightly , it
made in the light of the moon . It all boils down , wr
 surely under a crescent moon , thrilled by ice-white
sun , the seasons of the moon ? Home , alone , Jay pla
m is dazzling snow , the moon has risen full and cold
un and the temple of the moon , driving out of the hug
 in the dark and now the moon rises , full and amber a
bird on the shape of the moon over the trees in front
 But I could n't see the moon or the stars , only the
rning , with a sliver of moon hanging among the stars
 they love the sun , the moon and the stars . None of
the light of an enormous moon . The plash of flowing w
man 's first step on the moon ; various exhibits , aer
 the inevitable piece of moon rock . Housing The Airsh
oud obscured part of the moon . The Allied guns behind
```

(Baroni, 2014)

Distributional semantic models

- count how many times a target word occurs with different contexts

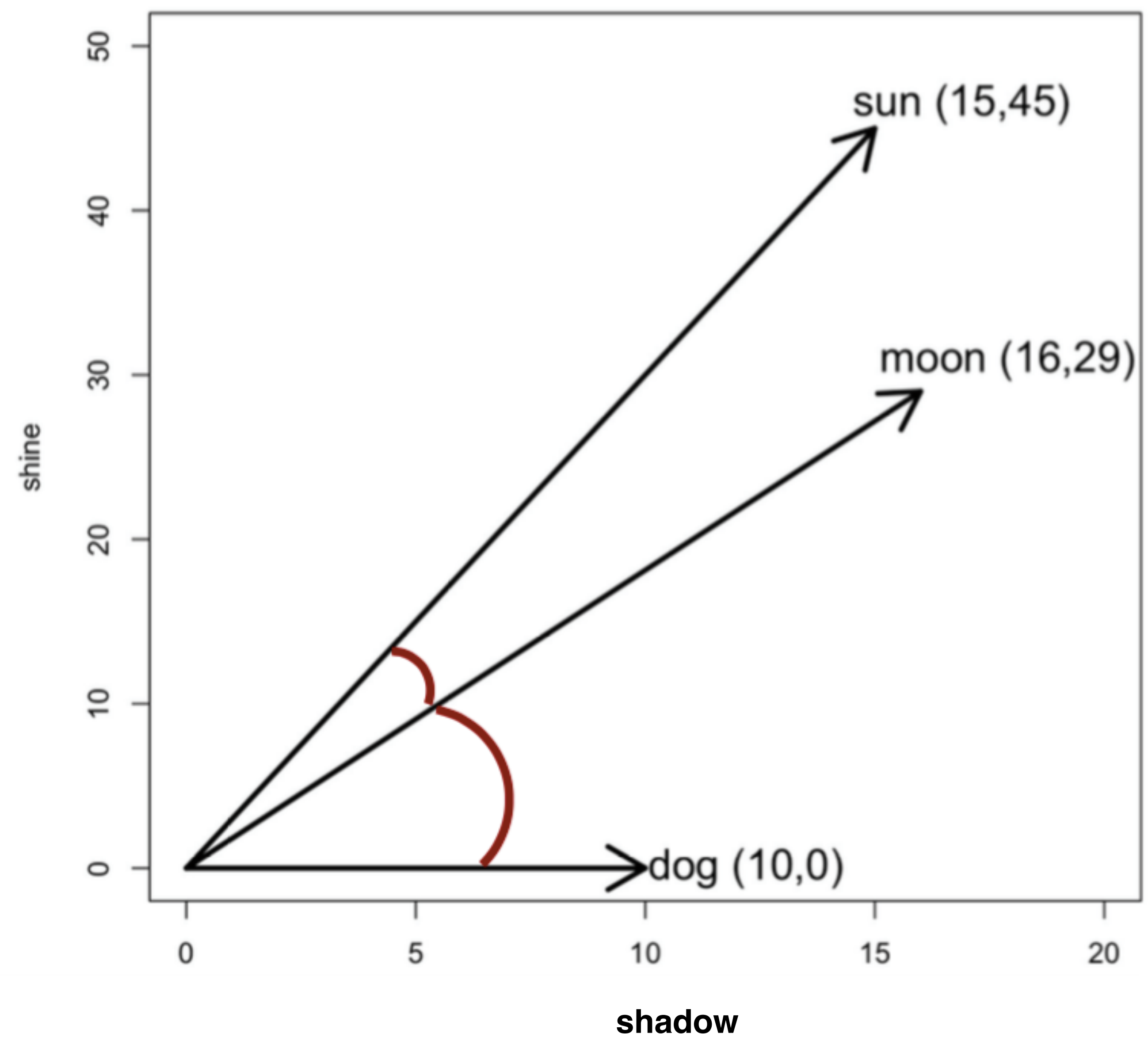
- context
  - document
  - sentence
  - fixed window around the word
  - syntactic dependencies

Co-occurrence matrix

	planet	night	full	shadow	shine	crescent
moon	10	22	43	16	29	12
sun	14	10	4	15	45	0
dog	0	4	2	10	0	0

Measuring similarity in meaning

	shadow	shine
moon	16	29
sun	15	45
dog	10	0



Cosine, Euclidean distance, Dice, Jaccard, …

Unsupervised WSI and WSD

Clustering for Sense Induction (Schütze, 1992;1998)

1. For each token w_i of word w in a corpus, compute a context vector of bag-of-words features $\vec{c}$

2. Use a clustering algorithm to group these vectors into a predefined number of clusters. Each cluster defines a sense of w.

3. Compute the vector centroid of each cluster. Each vector centroid is a sense vector $\vec{s_j}$ representing that sense of w.

Unsupervised WSI

A	S1	The **mouse** is also used a lot in scientific research though it is not an easy animal to examine
B	S2	Some **mouse** designs work like a joystick and may help. You can also use a touchpad ...
C	S1	**Mice** are great animals for several reasons. They are small, inexpensive,...
D	S2	I've been trying to install a new **mouse** on my touchpad but I have not succeeded yet...

(Klapaftis and Manandhar, 2013)

Unsupervised WSI and WSD

- agglomerative clustering
- *k*-means clustering
- topic modeling (Latent Dirichlet Allocation (LDA)) (Lau et al., 2012)
- spectral clustering (Zelnik-Manor and Perona, 2004)
- non-negative matrix factorization (Liu et al., 2013)

Unsupervised WSI and WSD

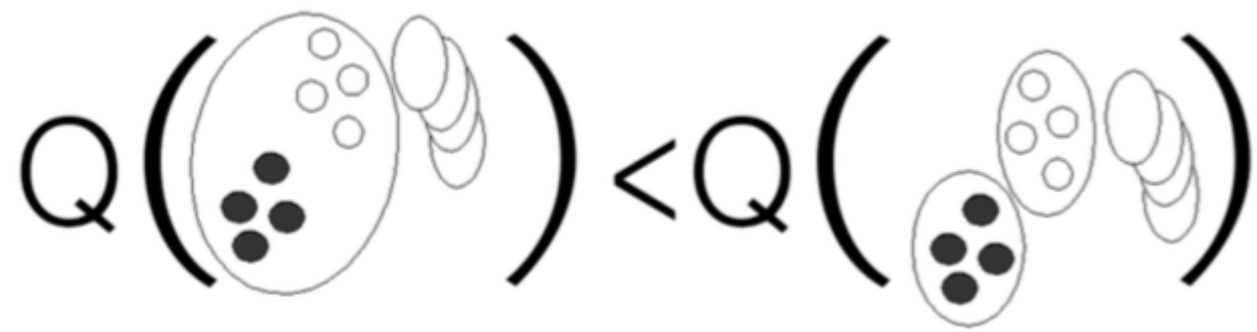
Word Sense Disambiguation

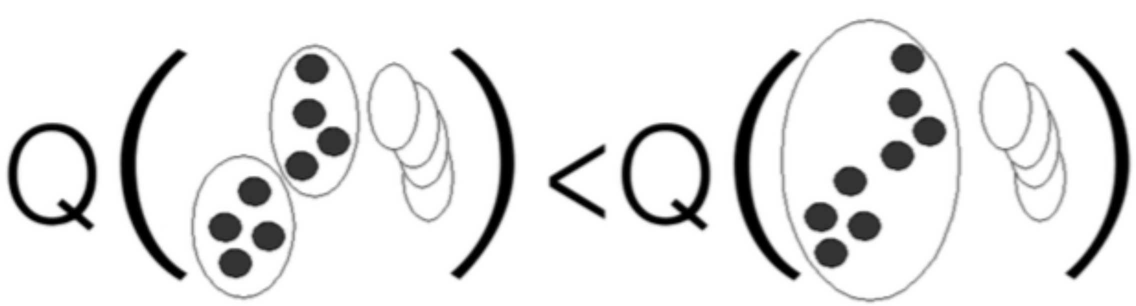
1. Compute a context vector $\vec{c}$ for an instance t of w
2. Retrieve all sense vectors $\vec{s_j}$ for w
3. Assign t to the sense represented by the sense vector $\vec{s_j}$ that is closest to t.

WSI - Intrinsic Evaluation

Comparison to a set of gold standard clusters

Homogeneity: clusters should not mix items belonging to different categories





Completeness: items belonging to the same category should be grouped in the same cluster

(Amigo et al., 2009)

WSI - Intrinsic Evaluation

- map the automatically derived clusters into a gold-standard set of senses(e.g. WordNet) by choosing the sense that has the most overlap with the cluster
- evaluate in a standard WSD task

(Manandhar et al., 2010; Jurgens and Klapaftis, 2013)

Part 4
Word Similarity

Word Similarity

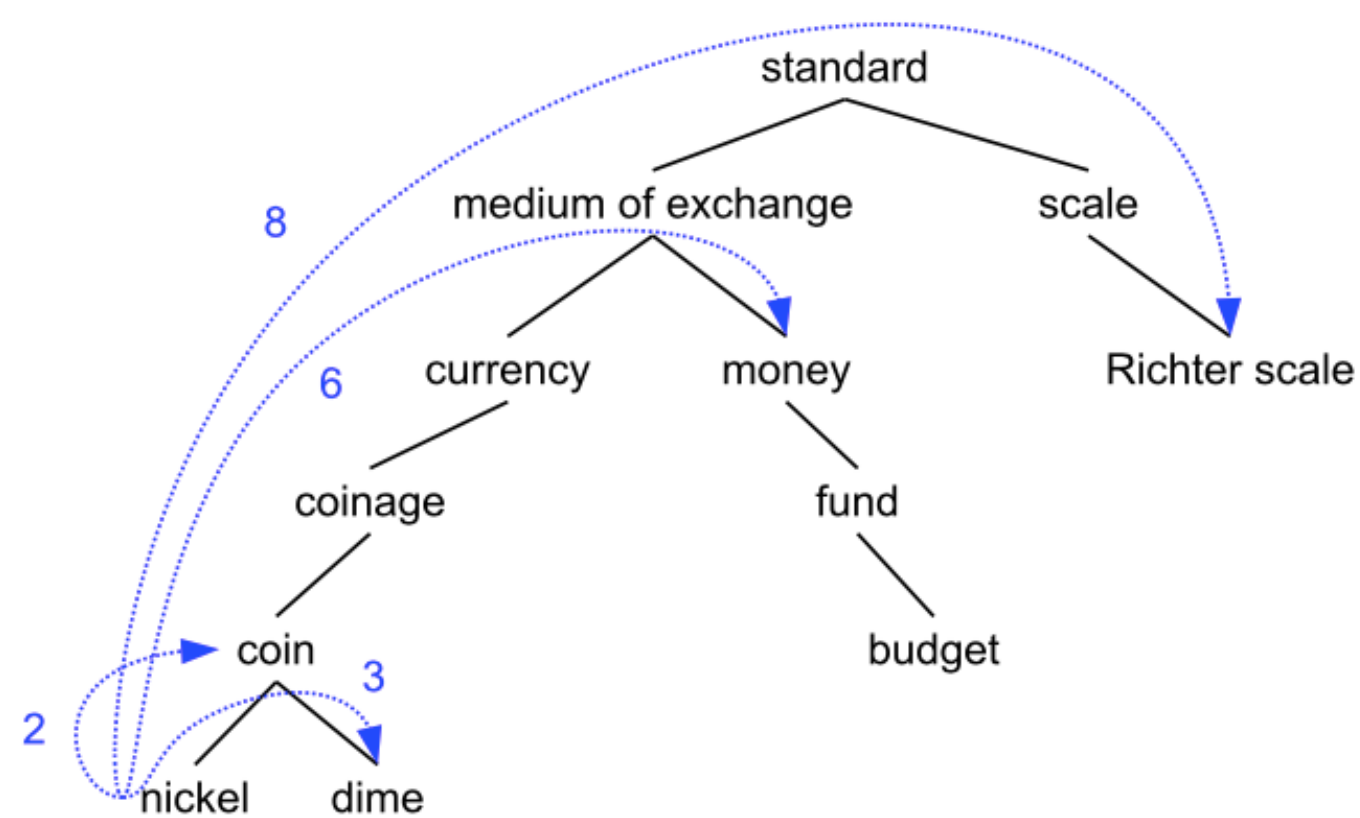
- two words are more/less similar if they share more/less features of meaning
- similarity and distance are relations between **word senses**
- applications
  - Natural Language Understanding
  - information retrieval or question answering
  - summarization, generation, and MT
  - language modeling (clustering for class-based models)

Word Similarity vs Relatedness

- **similar words**: near-synonyms or roughly substitutable in context
- **related words**: larger set of potential relationships
  - antonyms: *heavy - light*
  - thematically related words: *car - gasoline*
- word similarity is a subcase of word relatedness

Thesaurus-based similarity

Words/senses more similar if there is a **shorter path** between them in the graph (Quillian, 1969)



pathlen(c_1,c_2) = 1 + the number of edges in the shortest path between sense nodes c_1 and c_2

Thesaurus-based similarity

Path-based similarity (Leacock and Chodorow, 1998)

$$sim_{path}(c_1,c_2) = \frac{1}{pathlen(c_1,c_2)}$$

Word similarity ≈pair of senses with max sense similarity (Resnik,1995)

$$wordsim(w_1,w_2) = \max_{\substack{c_1 \in senses(w_1) \\ c_2 \in senses(w_2)}} sim(c_1,c_2)$$

Normalize using depth in the hierarchy (Wu and Palmer, 1994)

- links deep/higher up in the hierarchy represent a narrow/wider distance

Information-content similarity

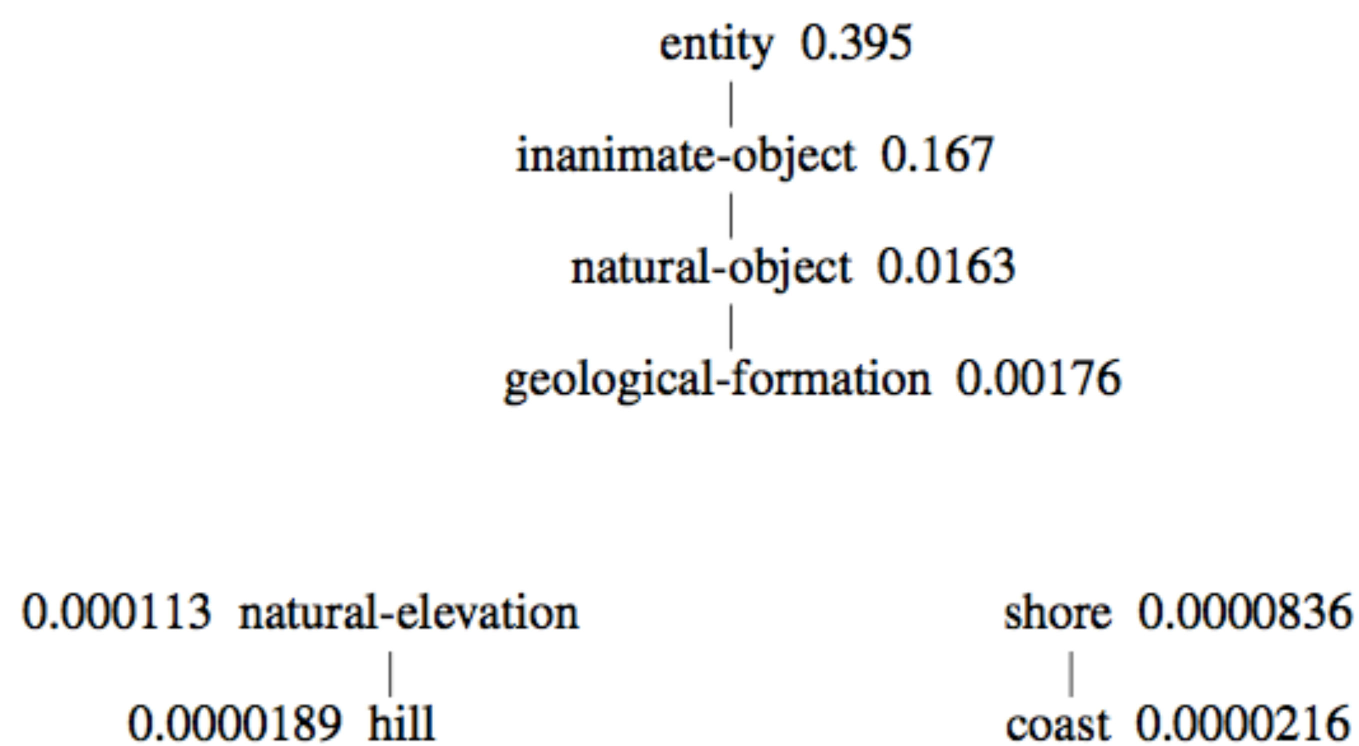
- Combine thesaurus structure + probabilistic information (Resnik, 1995)
- ***P(c)***: probability that a randomly selected word in a corpus is an instance of concept *c*
- ***P(root)*** = 1 (any word is subsumed by the root concept)
- the lower a concept in the hierarchy, the lower its probability

- each word in a corpus counts as an occurrence of each concept that contains it
- **words(*c*)**: the set of words subsumed by concept ***c***
- ***N***: total number of words in the corpus that are present in the thesaurus

$$P(c) = \frac{\sum_{w \in words(c)} count(w)}{N}$$

Information-content similarity

Lin (1998): probabilities from the WSJ + San Jose Mercury corpus



Information-content similarity

- the more two words have in common, the more similar they are
- common information of two nodes: information content of their lowest common subsumer (LCS), i.e. the lowest node in the hierarchy that is a hypernym of both c_1 and c_2

$$sim_{resnik}(c_1, c_2) = -\log P(LCS(c_1, c_2))$$

Information-content similarity

Similarity between concepts A and B (Lin, 1998)

- the ratio between the information needed to state the commonality of A and B and that to fully describe what A and B are

$$sim_{Lin}(A,B) = \frac{common(A,B)}{description(A,B)}$$

Information in common between two concepts is twice the information in the *LCS*(*c1,c2*)

$$sim_{Lin}(c_1,c_2) = \frac{2x \log P(LCS(c_1,c_2))}{\log P(c_1) + \log P(c_2)}$$

Information-content similarity

entity 0.395

inanimate-object 0.167

natural-object 0.0163

geological-formation 0.00176

0.000113 natural-elevation

0.0000189 hill

shore 0.0000836

coast 0.0000216

$$sim_{Lin}(hill, coast) = \frac{2x \log P(geo\log ical - formation)}{\log P(hill) + \log P(coast)} = 0.59$$

Dictionary-based similarity

Extended Lesk measure: two concepts are similar if their glosses contain overlapping words (Banerjee and Pedersen, 2003)

drawing paper: paper that is specially prepared for use in drafting

decal: the art of transferring designs from specially prepared paper to a wood or glass or metal surface

for each *n*-word phrase occurring in both glosses, add in a score of n^2 ($1^2 + 2^2 = 5$)

Dictionary-based similarity

Overlaps between glosses + glosses of hypernyms, hyponyms, meronyms, …

similarity(A,B) = overlap(gloss(A),gloss(B))
+ overlap(gloss(hypo(A)),gloss(hypo(B)))
+ overlap(gloss(A),gloss(hypo(B)))
+ overlap(gloss(hypo(A)),gloss(B))

$$sim_{eLesk}(c_1,c_2) = \sum_{r,q \in RELS} overlap(gloss(r(c_1)), gloss(q(c_2)))$$

RELS: the set of possible WordNet relations whose glosses are compared

Similarity Measures Evaluation

Intrinsic Correlation between an algorithm's similarity scores and human-defined similarity ratings

Out of context

- the **RG-65** dataset (Rubenstein and Goodenough, 1965)
- the **MC-30** dataset (Miller and Charles, 1991)
- **WordSim-353** (Finkelstein et al., 2002)
- **SimLex-999**
- **TOEFL** dataset

In context

- Stanford Contextual Word Similarity (SCWS) dataset (Huang et al., 2012)

- SemEval semantic textual similarity task (Agirre et al., 2012; Agirre et al., 2015)

Extrinsic Question answering, web search result clustering, text simplification…

Summary

Word Senses: relations, nature, identification

Semantic Resources: WordNet

WSD: supervised, knowledge-based, semi-supervised

WSI: sense clustering, unsupervised WSD

Word Similarity: thesaurus/dictionary-based, information content

Questions?

Comments?