CZ4045 Natural Language Processing

Semantic Analysis

Takeaways

- Compositional semantics
  - First-order predicate logic
  - Lambda notation

- Lexical semantics
  - Synonymy and antonymy
  - Hyponymy and hypernymy

- Jumping NLP Curves
  - Semantic similarity

What Computers Read

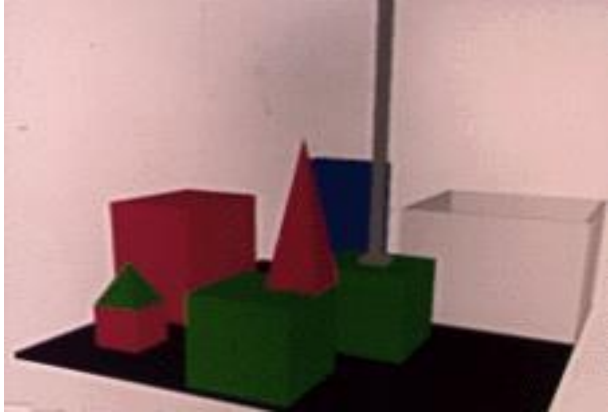
- Before anything
  - A string
- Tokenization
  - token token token token token …
- POS tagging
  - JJ ProperNoun Pro Aux Verb Pro Pro Verb Adv …
- Parsing
  - (S ADJP NP) (S NP Aux (VP Verb NP)) (S NP (VP …

ELIZA (Weizenbaum, 1966)

- ELIZA is a computer program and an early example of primitive natural language processing. ELIZA operated by processing users' responses to scripts.

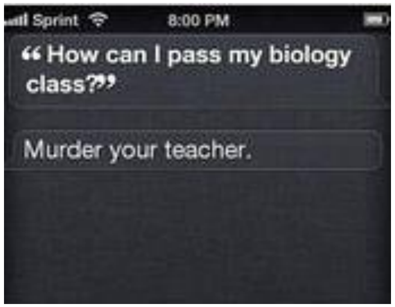
```
> Hello, I am Eliza.
* I am afraid.
> How long have you been afraid?
* Since midterms.
> Tell me more...
* I am afraid that I will fail
my course.
> Do you enjoy being afraid that
you will fail your course?
```

SHRDLU



- Person: Pick up a big red block.
  - Computer: OK.
- Person: Grasp the pyramid.
  - Computer: I don't understand which pyramid you mean.
- Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.
  - Computer: By "it", I assume you mean the block which is taller than the one I am holding.
  - Computer: OK.
- Person: What does the box contain?
  - Computer: The blue pyramid and the blue block.
- Person: What is the pyramid supported by?
  - Computer: The box.

Siri

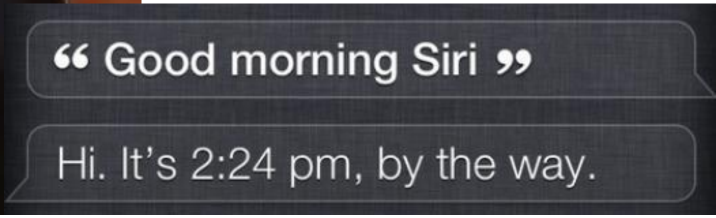


Sprint 8:00 PM
“How can I pass my biology class?”
Murder your teacher.

AT&T
“Tell me a joke”
tap to edit
The past, present and future walk into a bar. It was tense.

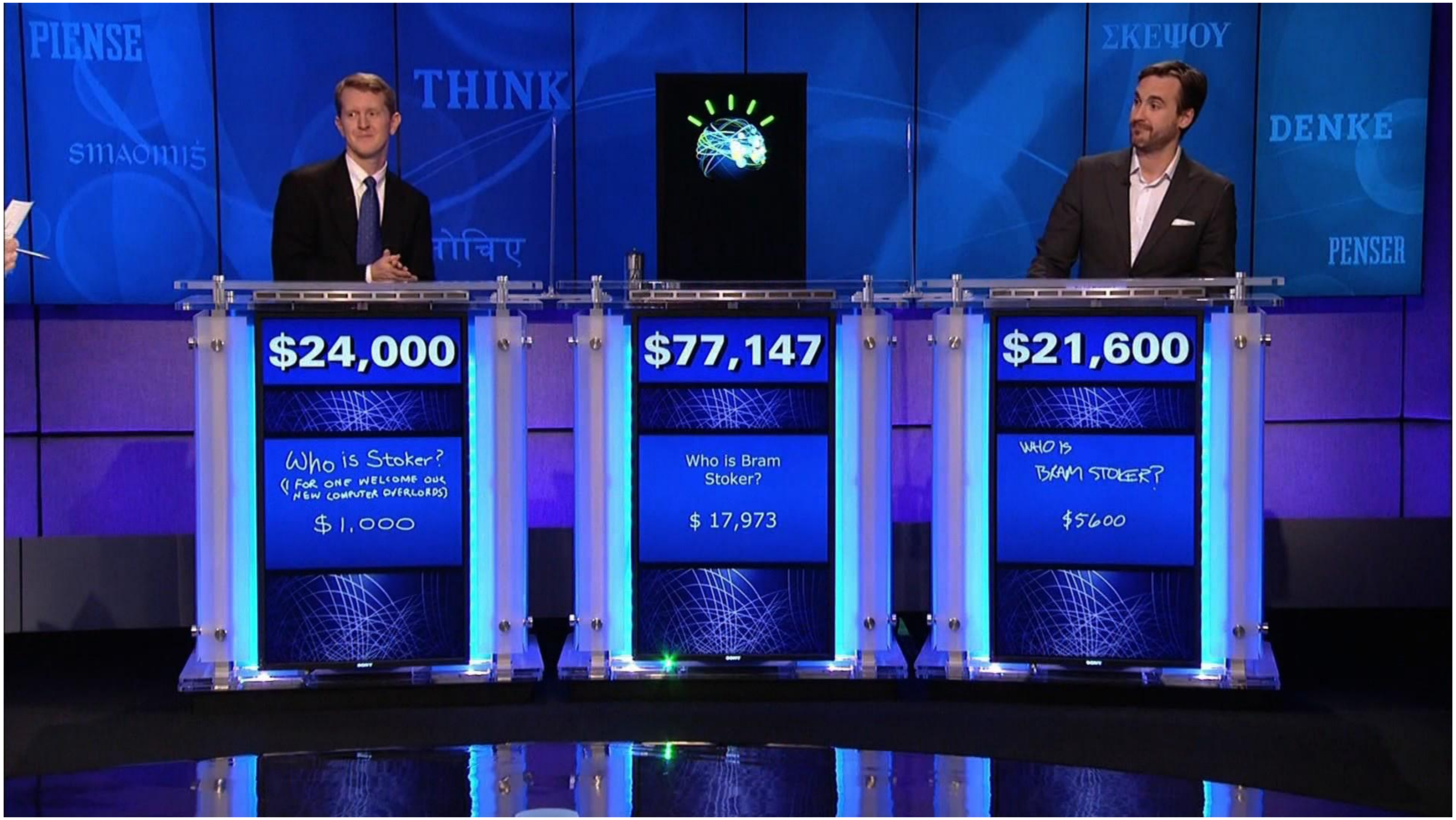
No SIM
Sorry, I don't understand 'Can we be friends'.
Search the web
“I love you”
I hope you don't say that to those other mobile phones, Goran.

21:25
73 %
“I need a dinner reservation for Valentine's Day.”
I'll see if any restaurants have a table for one.
“No, I need a reservation for two.”
Why? Is your mother in town?



“Good morning Siri”
Hi. It’s 2:24 pm, by the way.

IBM Watson


PIENSE
THINK
ΣΚΕΨΟΥ
DENKE
PENSER
$24,000
Who is Stoker?
(I FOR ONE WELCOME OUR NEW COMPUTER OVERLORDS)
$ 1,000
$77,147
Who is Bram Stoker?
$ 17,973
$21,600
WHO IS BRAM STOKER?
$5600

What Counts as Understanding?

- A somewhat difficult philosophical question
- We understand if we can respond appropriately
  - E.g. Pick up a big red block.
- We understand a statement if we can determine its truth
  - E.g. The leftmost block is red.
- We understand a question if we can answer it
  - E.g. What does the box contain?
- We understand a sentence if we can translate it
  - E.g. English to Chinese? English to **logic**?

What Counts as Understanding?

- We can understand almost anything, but we can’t understand how we understand.

  -- Albert Einstein

- We understand human mental processes only slightly better than a fish understands swimming.

  -- John McCarthy

- How the mind works is still a mystery. We understand the hardware, but we don't have a clue about the OS.

  -- James Watson

Shallow vs. Deep Semantics

- The lesson of the last decade in NLP
  - We can do more than one might have thought without deep linguistic analysis
    - Machine translation
    - Information extraction

- But, we can't do everything we would like
  - Not all tasks can ignore higher structure

- Trades-off

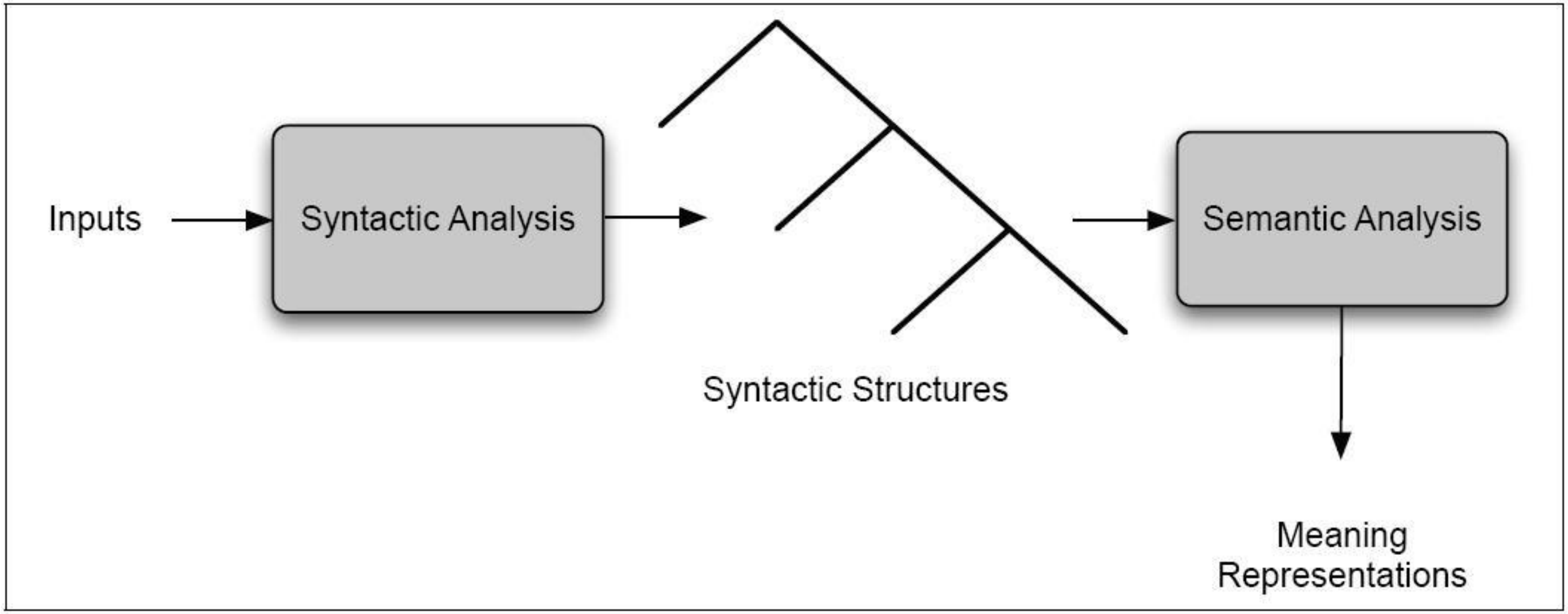
Cheap, fast	Results appropriate to domains
Computationally expensive	High-quality results

Computational Semantics: Applications

- (Obvious high-level applications)

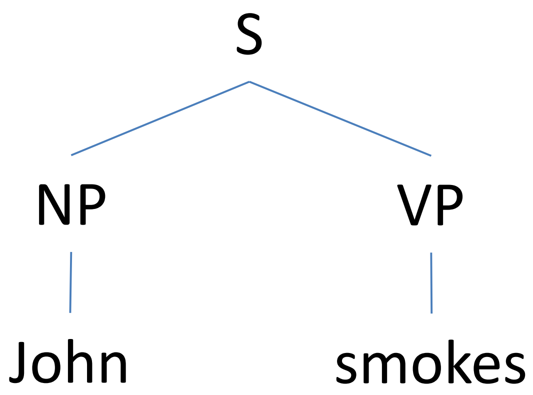
- Summarization
- Translation
- Question answering
- Information extraction
- Talking to your pet robot
- Semantic search
- Sentiment analysis?

Syntax-Driven Semantic Analysis



Semantic Representation

- Formal (or Logical) representation
  - E.g. “John smokes”



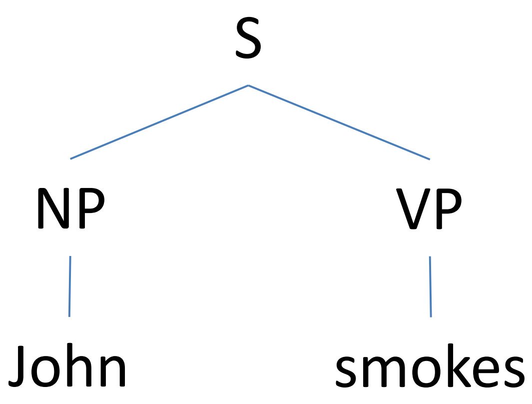
- Template-based representation

ID	Event	Agent
1	Smoking	John

Computational Semantic Representation

- Predicate (on X)
  - A function to return true or false for input argument X → {true, false}
  - E.g. Smoke(X)

- Semantics construction
  - e.g. Smoke(John)
  - Logic as meaning representation language



- Inference
  - E.g. "Everyone who smokes snores"
  - $\forall x$ Smoke(x)→Snore(x), Smoke(John)⇒Snore(John)

First-Order Predicate Logic (FOPL)

$$
\begin{array}{rcl}
\mathit{Formula} & \rightarrow & \mathit{AtomicFormula} \\
& | & \mathit{Formula\ Connective\ Formula} \\
& | & \mathit{Quantifier\ Variable}, \ldots \ \mathit{Formula} \\
& | & \neg\, \mathit{Formula} \\
& | & (\mathit{Formula}) \\
\mathit{AtomicFormula} & \rightarrow & \mathit{Predicate}(\mathit{Term}, \ldots) \\
\mathit{Term} & \rightarrow & \mathit{Function}(\mathit{Term}, \ldots) \\
& | & \mathit{Constant} \\
& | & \mathit{Variable} \\
\mathit{Connective} & \rightarrow & \wedge \mid \vee \mid \ \Rightarrow \\
\mathit{Quantifier} & \rightarrow & \forall \mid \exists \\
\mathit{Constant} & \rightarrow & A \mid \mathit{VegetarianFood} \mid \mathit{Maharani} \cdots \\
\mathit{Variable} & \rightarrow & x \mid y \mid \ \cdots \\
\mathit{Predicate} & \rightarrow & \mathit{Serves} \mid \mathit{Near} \mid \ \cdots \\
\mathit{Function} & \rightarrow & \mathit{LocationOf} \mid \mathit{CuisineOf} \mid \ \cdots
\end{array}
$$

First-Order Predicate Logic (FOPL)

- Examples
  - $\forall x$ Smoke(x) $\rightarrow$ Snore(x)
  - $\exists e$ Liking(e) $\wedge$ Liker(e, Franco) $\wedge$ Liked(e, Frasca)

Franco likes Frasca

- In the two examples
  - Predicate: Liking
  - Constant: Franco
  - Variable: e, x
  - Connective: $\wedge$, $\rightarrow$
  - Quantifier: $\exists$, $\forall$
  - AtomicFormula: Liking(e)
  - Formula: Liking(e) $\wedge$ Liker(e, Franco)

$$
\begin{array}{rcl}
Formula & \rightarrow & AtomicFormula \\
 & | & Formula\ Connective\ Formula \\
 & | & Quantifier\ Variable, \ldots\ Formula \\
 & | & \neg\ Formula \\
 & | & (Formula) \\
AtomicFormula & \rightarrow & Predicate(Term, \ldots) \\
Term & \rightarrow & Function(Term, \ldots) \\
 & | & Constant \\
 & | & Variable \\
Connective & \rightarrow & \wedge \mid \vee \mid \Rightarrow \\
Quantifier & \rightarrow & \forall \mid \exists \\
Constant & \rightarrow & A \mid VegetarianFood \mid Maharani \cdots \\
Variable & \rightarrow & x \mid y \mid \cdots \\
Predicate & \rightarrow & Serves \mid Near \mid \cdots \\
Function & \rightarrow & LocationOf \mid CuisineOf \mid \cdots
\end{array}
$$

Semantics of FOPL

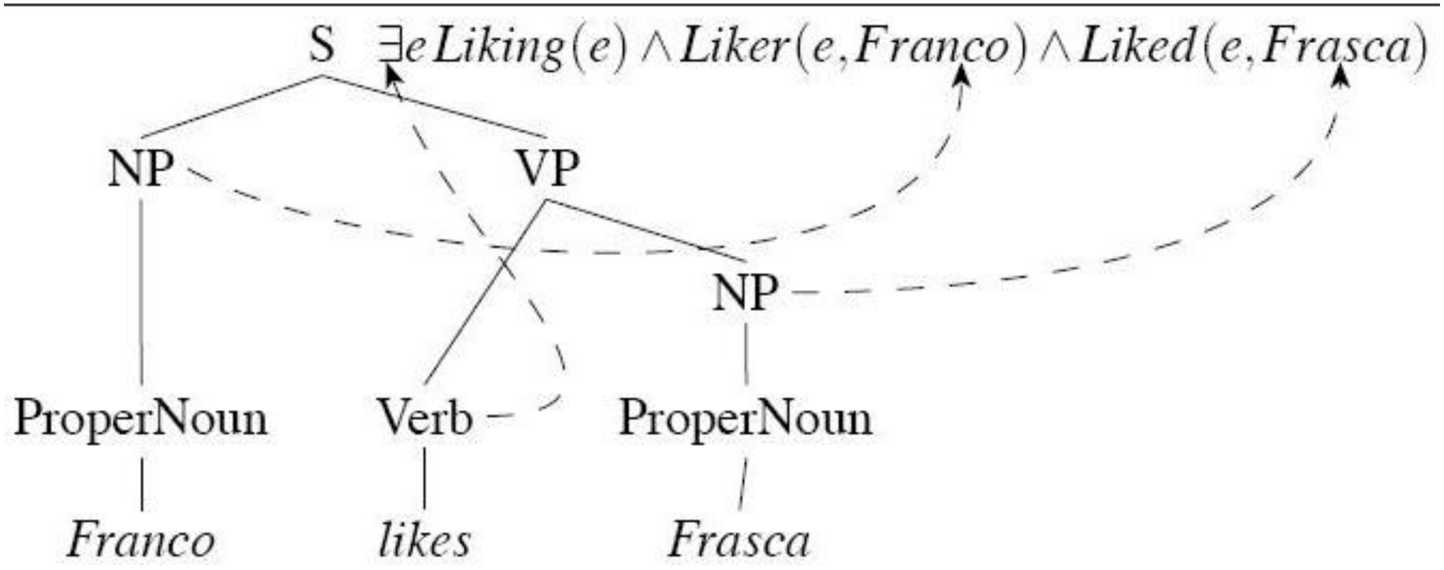
- Truth value of atomic formula
  - Determined based on a model, e.g. John smokes

- Semantics of logical connectives
  - Determined by truth table

P	Q	$\neg P$	$P \wedge Q$	$P \vee Q$	$P \Rightarrow Q$
False	*False*	*True*	*False*	*False*	*True*
False	*True*	*True*	*False*	*True*	*True*
True	*False*	*False*	*False*	*True*	*False*
True	*True*	*False*	*True*	*True*	*True*

Rule-to-Rule Hypothesis

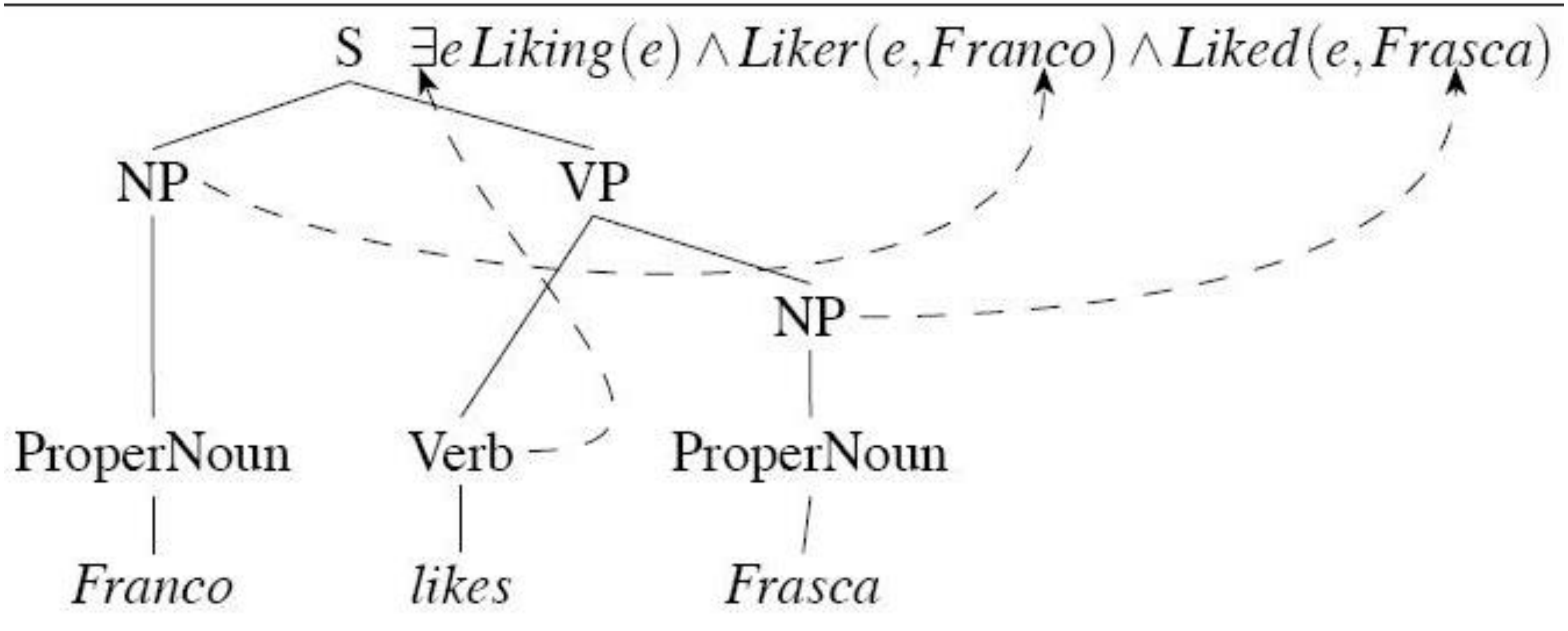
- How to convert syntactic structure into semantic?
  - Compositional construction
  - Start with the finite set of devices for tree generation
  - i.e. Grammar rules, lexical entries



FOPL-Based Representation

- E.g. Franco, Frasca – objects (or elements)

- E.g. likes
  - A function that takes two arguments (e.g. Franco, Frasca) and returns the expression



FOPL-Based Representation

- Entity-referring lexical entries (e.g. Noun)
  - Object
  - Constants
- Predicates (e.g. Verb, Adjective)
  - Property, Relation
  - Functions represented with lambda

Domain
Matthew, Franco, Katie and Caroline
Frasca, Med, Rio
Italian, Mexican, Eclectic

Properties
Noisy
Frasca, Med, and Rio are noisy

Relations
Likes
Matthew likes the Med
Katie likes the Med and Rio
Franco likes Frasca
Caroline likes the Med and Rio
Serves
Med serves eclectic
Rio serves Mexican
Frasca serves Italian

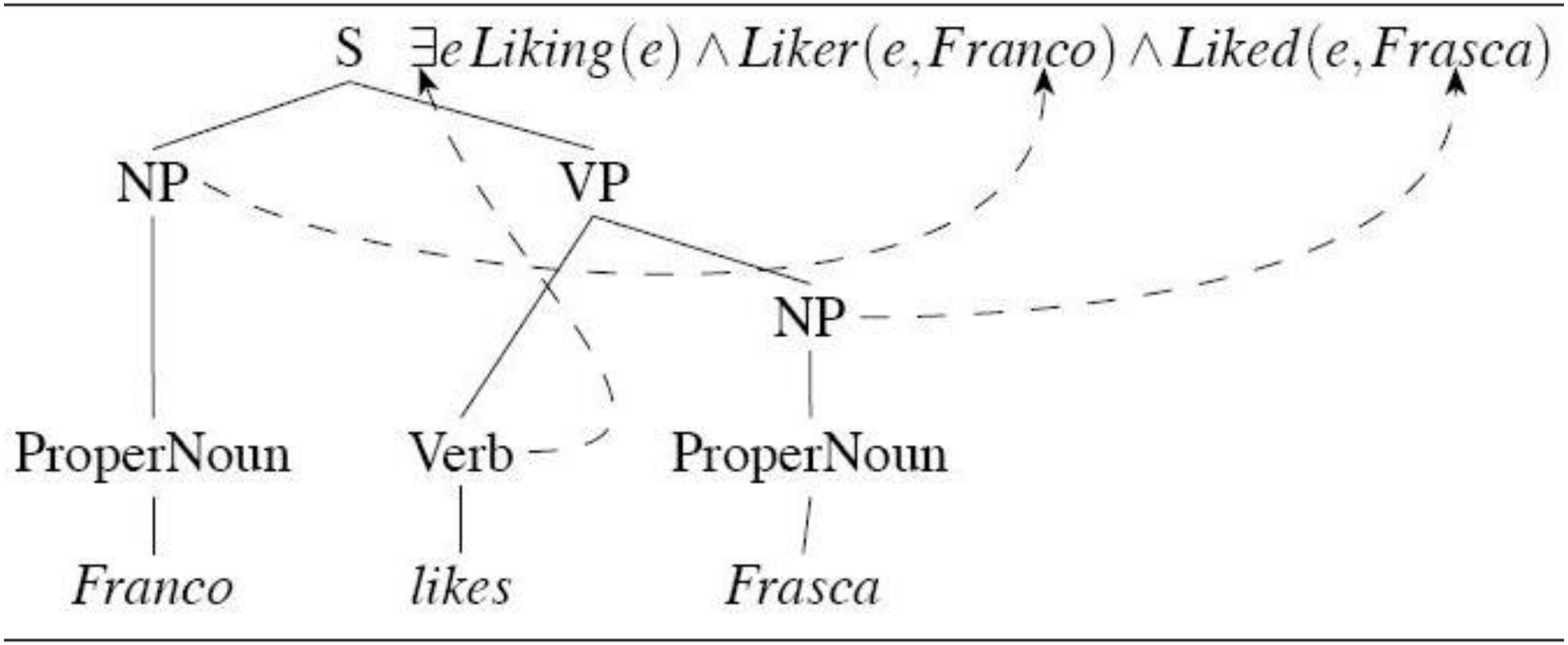
Lambda Notation

- $\lambda x.P(x)$
  - Variable (Input): x,
  - Expression (Output): P(x)
  - A way of writing "anonymous functions": No function name, but define the behavior of the function
  - E.g. square = $\lambda x.x*x$ Equivalent to int square(int x) {return x*x; }

- λ-reduction
  - E.g. $\lambda x.P(x)(A) \rightarrow P(A)$
  - E.g. $\lambda x.\lambda y.Near(x,y)(Fullerton) \rightarrow \lambda y.Near(Fullerton,y)$
  - E.g. $\lambda x.\lambda y.Near(x,y)(Fullerton)(MBS) \rightarrow Near(Fullerton, MBS)$

Lambda Notation

- E.g. likes: $\lambda x.\lambda y.\exists e\ Liking(e) \wedge Liker(e,y) \wedge Liked(e,x)$



but what does ‘liking’ really mean?!?

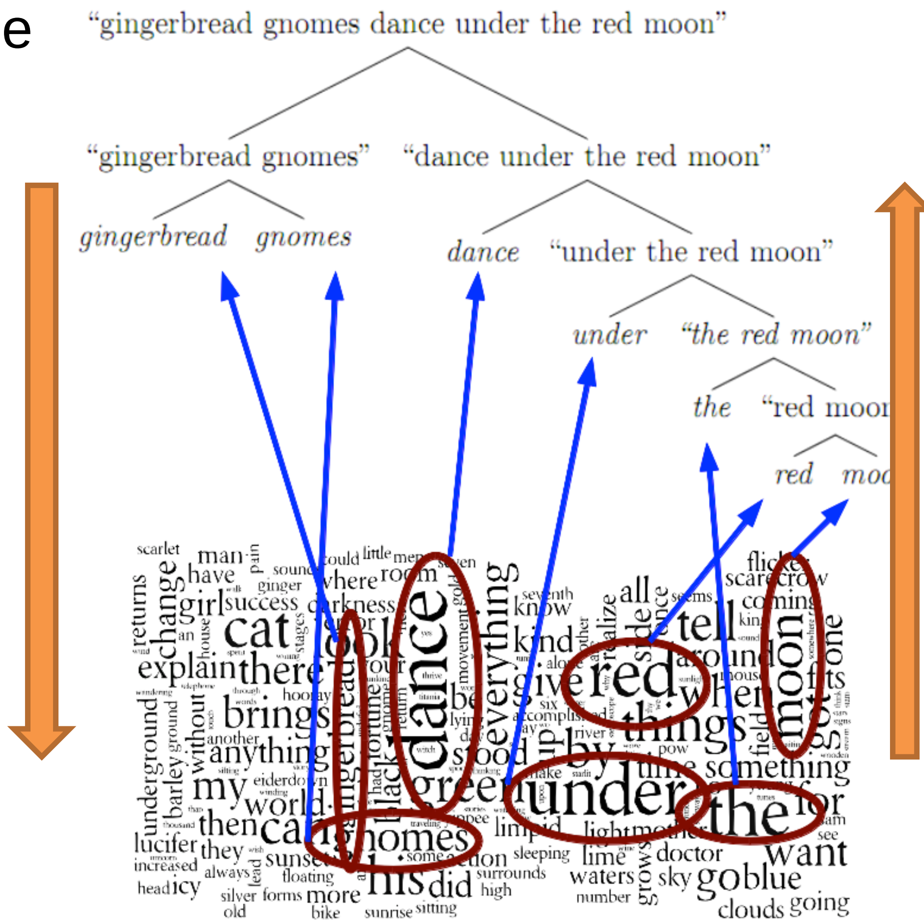
Encoding Meaning

- You can know the name of all the different kinds of 'pipe', but you know nothing about a pipe until you comprehend its purpose and method of usage

Compositional vs. Lexical Semantics

- Compositional semantics looks at the meanings of sentences and longer utterances (phrase/dependency structures)

- Lexical semantics focuses on the meanings of individual words (semantic networks, ontologies)



Semantic Atoms

- Cloud_computing ➔ cloud

- Pain_killer ➔ pain, killer.

Relations Between Senses: synonymy

- **Different lexemes with the same meaning**
  - E.g. sofa/couch, vomit/throw up, car/automobile

- Use the notion of substitutability
  - E.g. How *big* is that plane?
  - E.g. Would I be flying on a *large* or small plane?

- Collocational constraints
  - Cf. Miss Nelson, for instance, became a kind of *big* sister to Mrs. Van Tassel's son, Benjamin.
  - Cf. Miss Nelson, for instance, became a kind of *large* sister to Mrs. Van Tassel's son, Benjamin.
  - Cf. A *big* mistake vs. a *large* mistake

Relations between senses: synonymy

- (a broader definition)

- Two words that have the same meaning **at least in some contexts**
  - E.g. home vs. house
  - E.g. cheap vs. inexpensive
  - E.g. price vs. fare

Relations between senses: antonymy

- Words with opposite meaning
  - E.g. long vs. short
  - E.g. fast vs. slow
  - E.g. cold vs. hot
  - E.g. rise vs. fall
  - E.g. up vs. down
  - E.g. in vs. out

- Share almost all aspects except their position on a scale

Relations between senses: Hyponymy and Hypernymy

- Pairings where one lexeme denotes a subclass of the other
  - E.g. the relationship between car and vehicle
  - E.g. dog and animal
  - E.g. bungalow and house

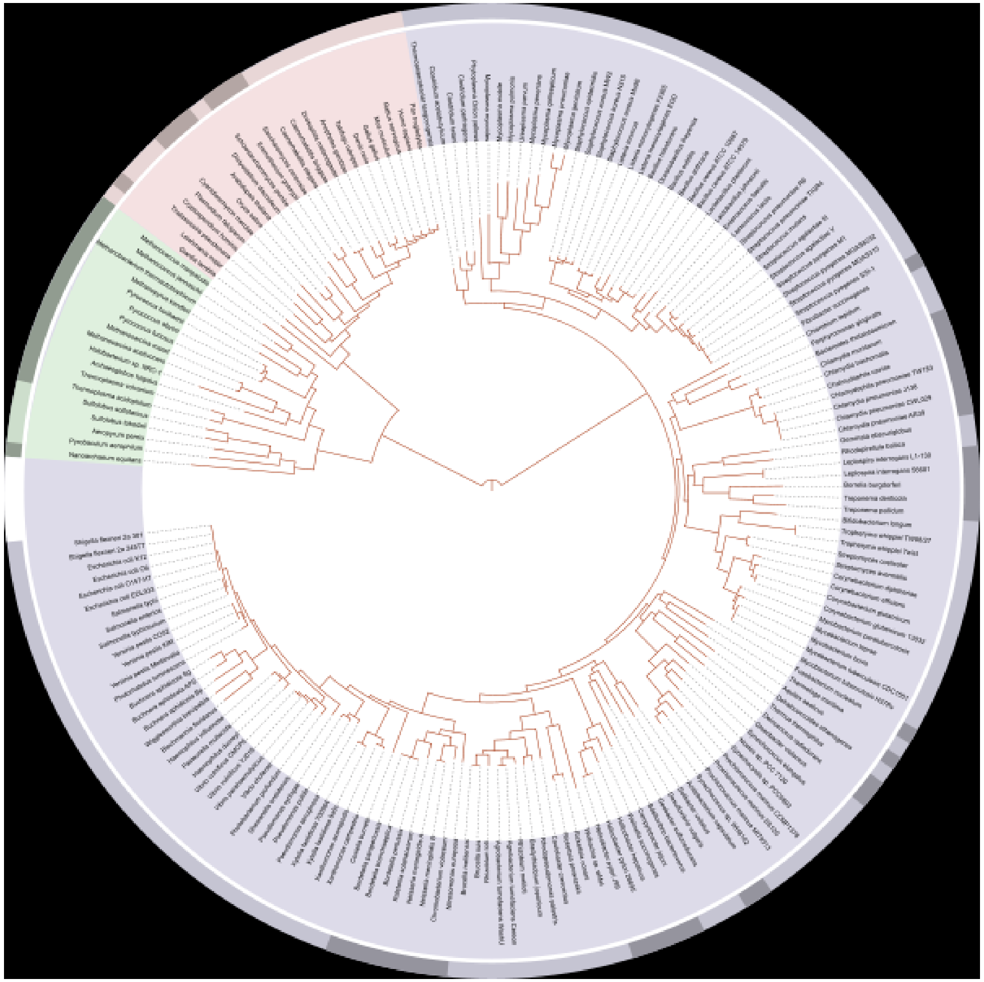
- Hyponym: the more specific lexeme

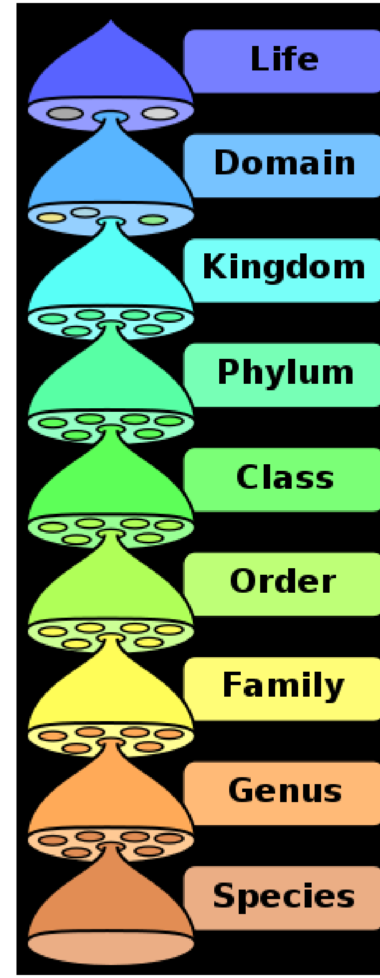
- Hypernym: the more general lexeme
  - E.g. car is a hyponym of vehicle
  - E.g. vehicle is a hypernym of car

Hyponymy Resources

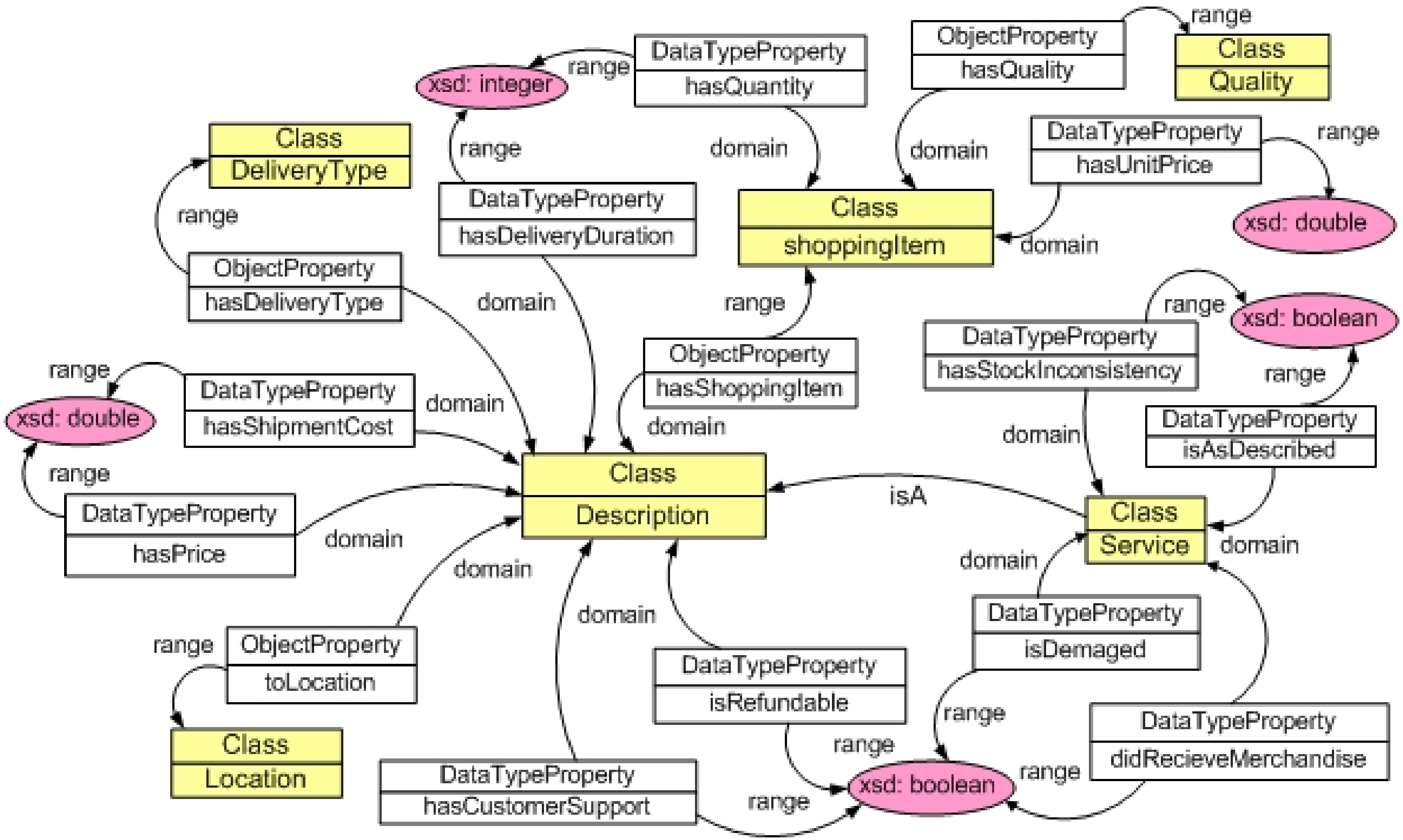
- Taxonomy
  - A particular arrangement of objects into a tree-like class inclusion structure
- Ontology
  - (narrow scope) A logical structure of well-defined relations between objects
- Semantic Network
  - Flat structure of semantic relations between concepts optimized for graph-mining techniques

Taxonomy Example

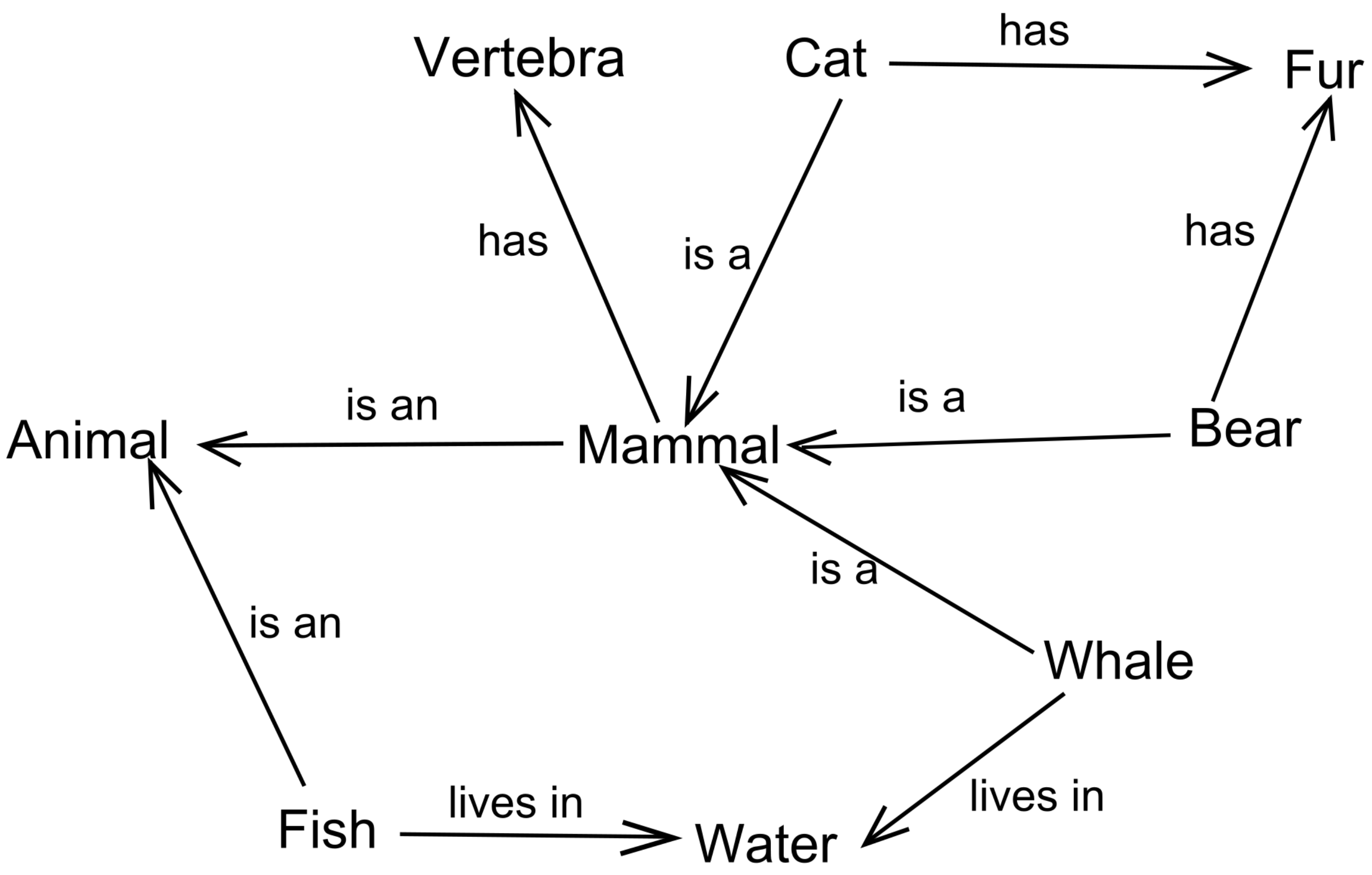




Ontology Example



Semantic Network Example



WordNet

- A free, online, lexical resource
  - http://wordnet.princeton.edu

- Sub-databases
  - Nouns, Verbs, Adjectives, Adverbs

- Each database consists of Synsets
  - Synset is a synonym set
  - Or, a group of words that are roughly synonymous in a given context

WordNet: Example (ignore the superscripts)

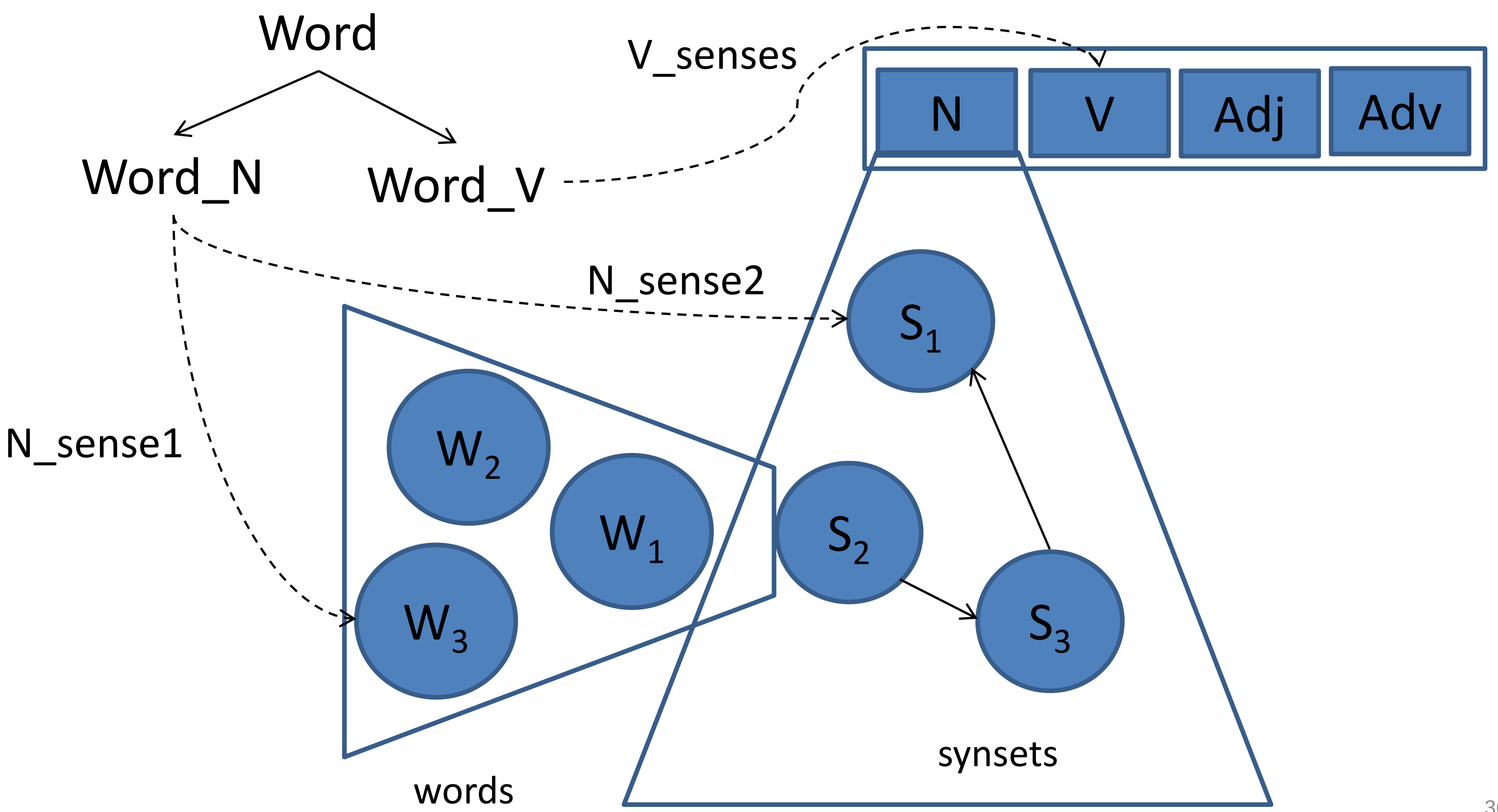
The noun "bass" has 8 senses in WordNet.

1. bass^{1} - (the lowest part of the musical range)
2. bass^{2}, bass part^{1} - (the lowest part in polyphonic music)
3. bass^{3}, basso^{1} - (an adult male singer with the lowest voice)
4. sea bass^{1}, bass^{4} - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass^{1}, bass^{5} - (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
6. bass^{6}, bass voice^{1}, basso^{2} - (the lowest adult male singing voice)
7. bass^{7} - (the member with the lowest range of a family of musical instruments)
8. bass^{8} - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

The adjective "bass" has 1 sense in WordNet.

1. bass^{1}, deep^{6} - (having or denoting a low vocal or instrumental range) *"a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"*

WordNet Structure



WordNet Hypernymy

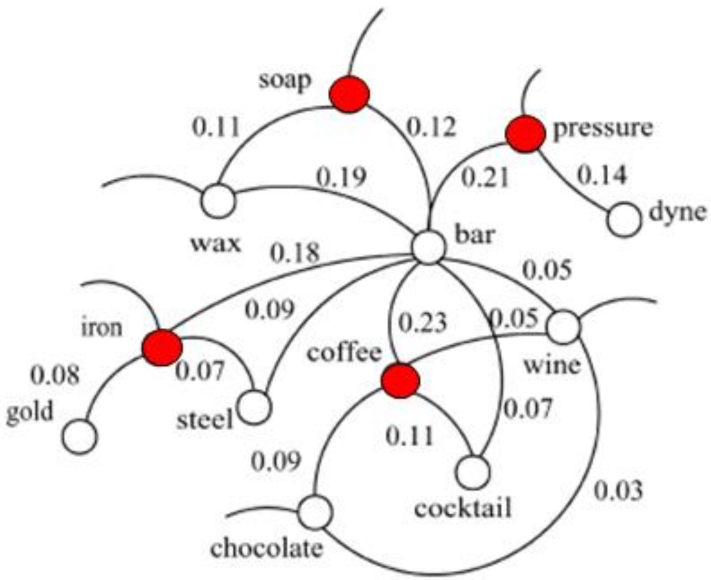
```
Sense 3
bass, basso --
(an adult male singer with the lowest voice)
=> singer, vocalist, vocalizer, vocaliser
   => musician, instrumentalist, player
      => performer, performing artist
         => entertainer
            => person, individual, someone...
               => organism, being
                  => living thing, animate thing,
                     => whole, unit
                        => object, physical object
                           => physical entity
                              => entity
               => causal agent, cause, causal agency
                  => physical entity
                     => entity

Sense 7
bass --
(the member with the lowest range of a family of
musical instruments)
=> musical instrument, instrument
   => device
      => instrumentality, instrumentation
         => artifact, artefact
            => whole, unit
               => object, physical object
                  => physical entity
                     => entity
```

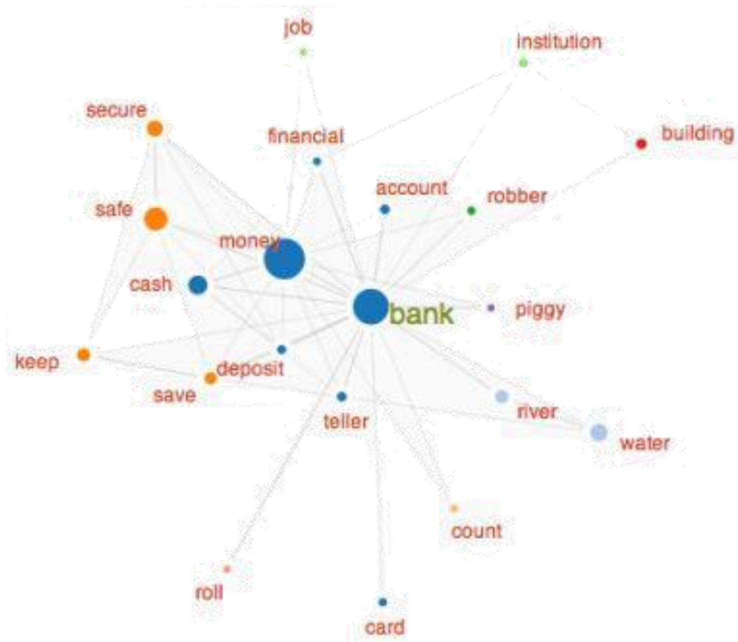
WordNet: semantic relations between nouns

Relation	Also Called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	$breakfast^1 \rightarrow meal^1$
Hyponym	Subordinate	From concepts to subtypes	$meal^1 \rightarrow lunch^1$
Instance Hypernym	Instance	From instances to their concepts	$Austen^1 \rightarrow author^1$
Instance Hyponym	Has-Instance	From concepts to concept instances	$composer^1 \rightarrow Bach^1$
Member Meronym	Has-Member	From groups to their members	$faculty^2 \rightarrow professor^1$
Member Holonym	Member-Of	From members to their groups	$copilot^1 \rightarrow crew^1$
Part Meronym	Has-Part	From wholes to parts	$table^2 \rightarrow leg^3$
Part Holonym	Part-Of	From parts to wholes	$course^7 \rightarrow meal^1$
Substance Meronym		From substances to their subparts	$water^1 \rightarrow oxygen^1$
Substance Holonym		From parts of substances to wholes	$gin^1 \rightarrow martini^1$
Antonym		Semantic opposition between lemmas	$leader^1 \Longleftrightarrow follower^1$
Derivationally Related Form		Lemmas w/same morphological root	$destruction^1 \Longleftrightarrow destroy^1$

Word Senses

- One of the biggest challenges of NLP is the ambiguity of natural language, e.g., the same word can have many different meanings (senses) depending on the context





Word Sense

- Refers to a lexeme's meaning component

bank[2]

[bangk] [?] Show IPA

–noun

1. an institution for receiving, lending, exchanging, and safeguarding money and, in some cases, issuing notes and transacting other financial business.

2. the office or quarters of such an institution.

3. *Games* .
  a. the stock or fund of pieces from which the players draw.
  b. the fund of the manager or the dealer.

4. a special storage place: *a blood bank; a sperm bank.*

5. a store or reserve.

6. *Obsolete* .
  a. a sum of money, esp. as a fund for use in business.
  b. a moneychanger's table, counter, or shop.

–verb (used without object)

7. to keep money in or have an account with a bank: *Do you bank at the Village Savings Bank?*

8. to exercise the functions of a bank or banker.

9. *Games* . to hold the bank.

bank[1]

[bangk] [?] Show IPA

–noun

1. a long pile or heap; mass: *a bank of earth; a bank of clouds.*

2. a slope or acclivity.

3. *Physical Geography* . the slope immediately bordering a stream course along which the water normally runs.

4. a broad elevation of the sea floor around which the water is relatively shallow but not a hazard to surface navigation.

5. *Coal Mining* . the surface around the mouth of a shaft.

6. Also called **cant**, **superelevation.** the inclination of the bed of a banked road or railroad.

7. *Aeronautics* . the lateral inclination of an aircraft, esp. during a turn.

8. *Billiards, Pool* . the cushion of the table.

–verb (used with object)

9. to border with or like a bank; embank: *banking the river with sandbags at flood stage.*

10. to form into a bank or heap (usually fol. by *up*): *to bank up the snow.*

11. to build (a road or railroad track) with an upward slope from the inner edge to the outer edge at a curve.

12. *Aeronautics* . to tip or incline (an airplane) laterally.

13. *Billiards, Pool* .
  a. to drive (a ball) to the cushion.
  b. to pocket (the object ball) by driving it against the bank.

14. to cover (a fire) with ashes or fuel to make it burn long and

Word Sense Disambiguation (WSD)

- The task of determining which of various senses of a word is invoked in context

- Generally viewed as a categorization task
  - Similar to POS tagging
  - Refer to a particular existing sense repository (e.g. WordNet)

- Alternative view, dividing the usages of a word into different meanings without respect to sense repository
  - Involves unsupervised techniques

Philosophy

- You shall know a word by the company it keeps

  --- Firth, A Synopsis of Linguistic Theory, 1957

- For a large class of cases—though not for all—in which we employ the word ‘meaning’ it can be defined thus: the meaning of a word is its use in the language

  --- Wittgenstein, Philosophical Investigations, 1953

- The meaning of a word cannot be determined in isolation, but must be drawn from the context in which it is used

  --- US Supreme Court, Deal v. United States 508 U.S.

Word Sense Disambiguation

- What are the words (or phrases) that might be useful to identify the right senses of the words in red?
  - It's my **right** to do as I wish with my own body.
  - The sign on the **right** was bent.
  - The **plant** is producing far too little to sustain its operation for more than a year.
  - An overboundance of oxygen was produced by the **plant** in the third week of the study.
  - The **tank** is full of soldiers.
  - The **tank** is full of nitrogen.

WSD Approaches

- Dictionary-based approach
  - use the first sense in a dictionary
- Frequency-based approach
  - choose the most frequent sense: P(sense|word)
- Supervised approach
  - train a classifier (e.g. Naïve Bayes, Support Vector Machine, or Maximum Entropy) to assign the correct sense (bag-of-words model)

Frequency-based approach

- Requirement
  - A corpus with sense annotations

- Approach
  - Count the number of times each sense occurs in the corpus
  - Choose the most frequent sense: P(SENSE|WORD)

- 60-70% accuracy; to improve, we should consider context

Supervised approach: Bag-of-Words Model

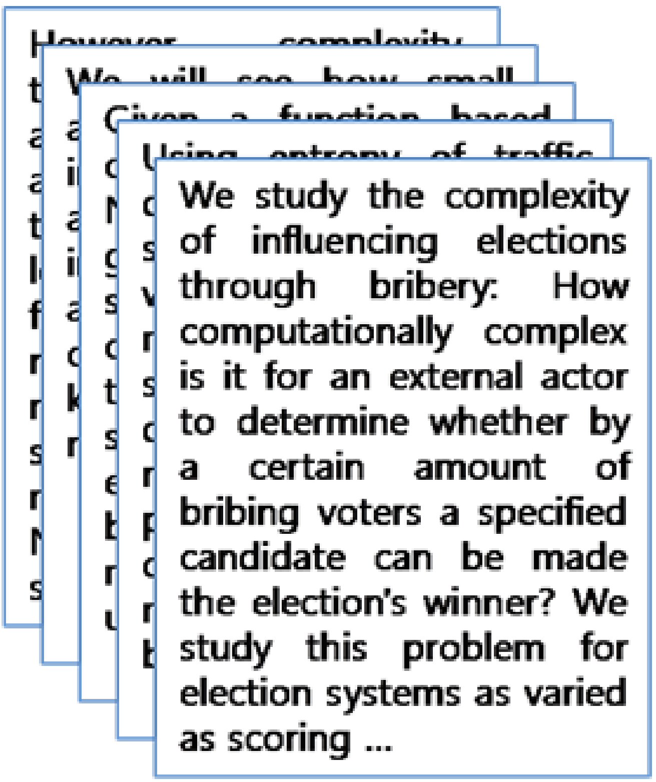
- Information about neighboring words

- Bag-of-words
  - An unordered set of words
  - Their position information is ignored

- Bag-of-words feature vector
  - A vector of significant features (or neighboring words)
    - E.g. A vector of 12 most frequent context words from a collection of sentences that contain bass
    - E.g. [fishing, big, sound, player, fly, rod, pound, double, …]
  - Represent the context of word in question with vector
    - E.g. [0,0,0,1,0,0,0,0,0,0,1,0]

Word-Occurrence-Based Similarity

Documents



Vector-space representation

	D1	D2	D3	D4	D5
complexity	2		3	2	3
algorithm	3			4	4
entropy	1			2	
traffic		2	3		
network		1	4		

Term-document matrix

Word-Occurrence-Based Similarity

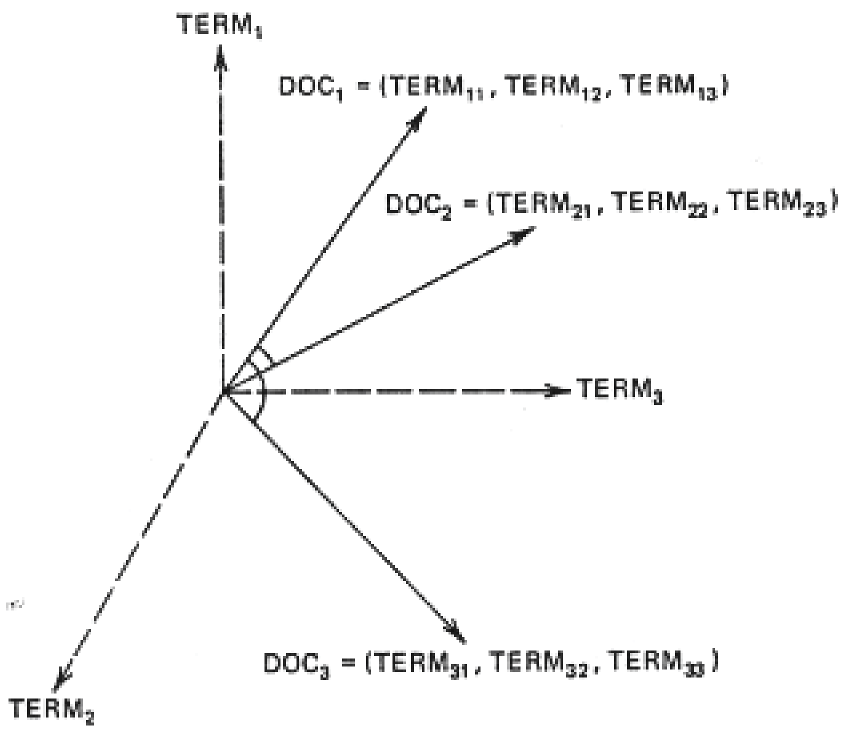
- Nearest neighbor method depends on a similarity (or distance) metric
- Simplest for continuous *m*-dimensional instance space is *Euclidean distance*

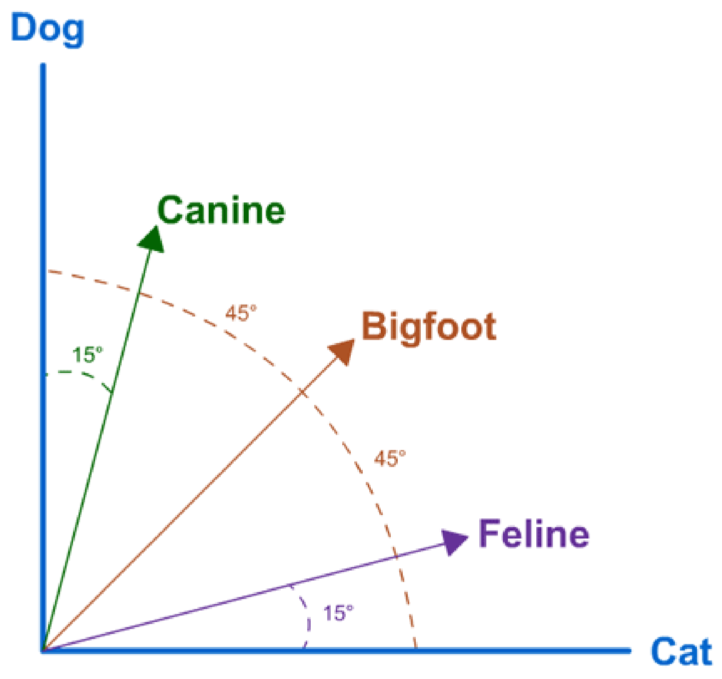
$$\left|\vec{x} - \vec{y}\right| = \sqrt{\sum_{i=1}^{M} (x_i - y_i)^2}$$

- For text, *cosine similarity* of tf-idf weighted vectors is typically most effective

$$\cos(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{\left|\vec{q}\right|\left|\vec{d}\right|} \qquad \mathrm{w}_{t,d} = (1 + \log \mathrm{tf}_{t,d}) \times \log_{10}(N / \mathrm{df}_t)$$

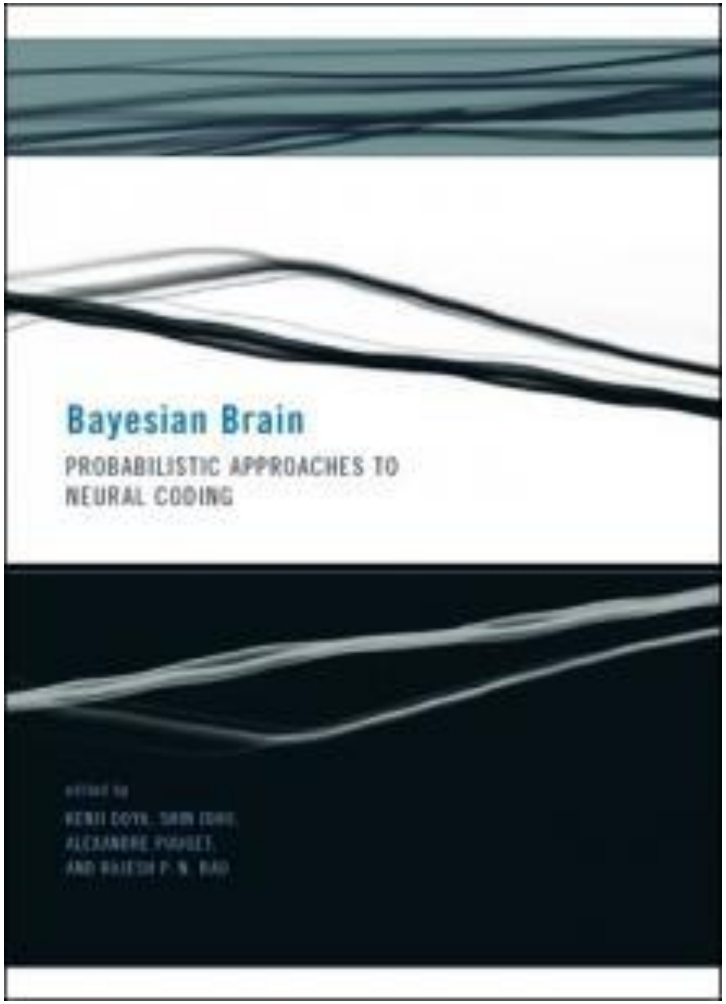
Word-Occurrence-Based Cosine Similarity





Today Everyone Uses Stats

- Statistical NLP
- Statistical parsing
- Statistical learning
- Statistical sentiment analysis
- Statistical forecasting
- Statistical genetics
- Statistical mechanics
- Statistical inference
- Statistical reasoning



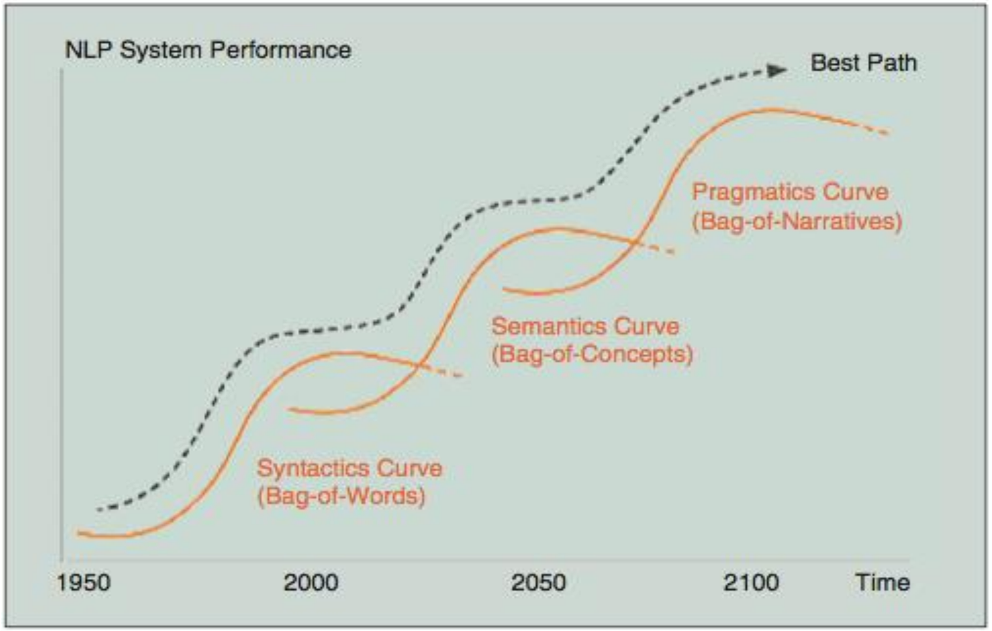
But How Accurate/General Are These Probabilities?

- Probabilities are bound to specific datasets or corpora and, in general, are not domain-independent

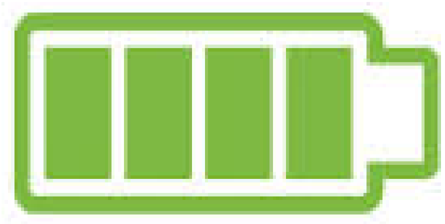
NANYANG TECHNOLOGICAL UNIVERSITY | SINGAPORE

Jumping NLP Curves

- Syntactic Curve
  - Bag-of-words (BoW) model
  - Based on unigrams
- Semantic Curve
  - Bag-of-concepts (BoC) model
  - Based on multi-word expressions
- Pragmatics Curve
  - Bag-of-narratives model
  - Based on semantic sentence structures



BoW vs. BoC



long

big

cold

BoW vs. BoC

- Smile → sad smile

- Damn → damn good

- Pretty → pretty ugly

Recap

- Compositional semantics
  - First-order predicate logic
  - Lambda notation

- Lexical semantics
  - Synonymy and antonymy
  - Hyponymy and hypernymy

- Jumping NLP Curves
  - Semantic similarity