BBM 495
LEXICAL SEMANTICS
WORD SENSE DISAMBIGUATION

LECTURER: BURCU CAN

2019-2020 SPRING

Outline

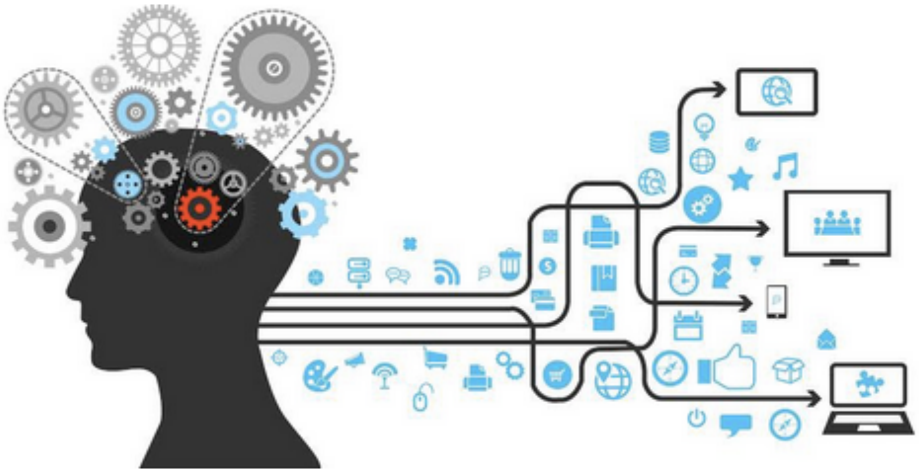
- So far we've focused on
  - Words (morphology, spelling correction, etc)
  - PoS tags of words
  - Syntactic structure (e.g. parse trees) of sentences

- Now comes the meaning…

"You're fired, eliminated, laid off, let go, downsized, terminated …"

Meaning

- The grand goal of artificial intelligence
  - machines that do not mindlessly process data
  - … but ultimately understand its meaning



Semantics

- To build our QA system we will need to deal with issues in **semantics**, i.e., meaning.

It comes in two flavors:

- Lexical Semantics: The meaning of individual words
- Compositional semantics: How the meaning of individual units combine to form the meaning of larger units

Lexical vs. Compositional Semantics

- LEXICAL
- man: An adult male human.
- dog: A domesticated carnivorous mammal

- COMPOSITIONAL
- Dog bites man. (*happens all the time; not too interesting)*
- Man bites dog. (*newsworthy)*

Example Question (1)

- Question
  - When was Ada Lovelace born?

- Text available to the machine
  - Ada Lovelace was born on December, 1815.

- This is easy.
  - just phrase a Google query properly: ”Ada Lovelace was born on *"
  - syntactic rules that convert questions into statements are straight-forward

Example Question (2)

- Question
  - What plants are native to Turkey?
- Text available to the machine
  - A new chemical plant was opened in Turkey.
- What is hard?
  - words may have different meanings (senses)
  - we need to be able to disambiguate between them

Example Question (3)

- Question
  - Where did Theresa May go on vacation?
- Text available to the machine
  - Theresa May spent her holiday in Cornwall
- What is hard?
  - words may have the same meaning (synonyms)
  - we need to be able to match them

Example Question (4)

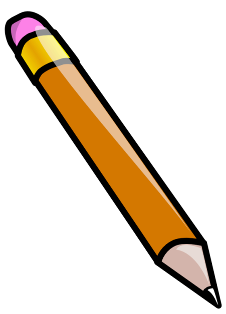
- Question
  - Which animals love to swim?
- Text available to the machine
  - Polar bears love to swim in the freezing waters of the Arctic.
- What is hard?
  - words can refer to a subset (hyponym) or superset (hypernym) of the concept referred to by another word
  - we need to have database of such A is-a B relationships, called an ontology

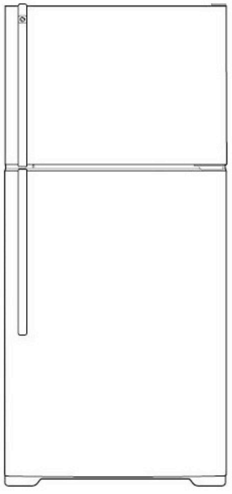
Example Question (5)

- Question
  - Did Poland reduce its carbon emissions since 1989?
- Text available to the machine
  - Due to the collapse of the industrial sector after the end of communism in 1989, all countries in Central Europe saw a fall in carbon emissions.
  - Poland is a country in Central Europe.
- What is hard?
  - we need to do inference
  - a problem for sentential, not lexical, semantics

Semantic Relationships

- Semantic relationships indicate a similarity in meaning between two words.
  - **“crayon” and “pencil.”**
  - **But not “pencil” and “refrigerator”, for example.**

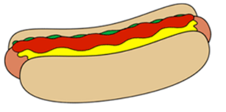


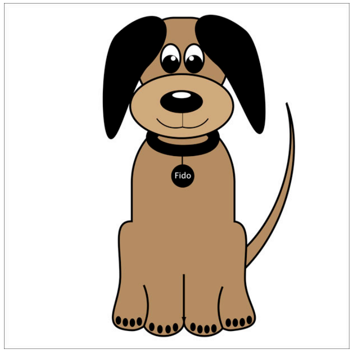


- So far, we have focused on the structure of language, not on what things *mean.*
- We have been doing natural language *processing*, but not natural language *understanding*.
- So what *is* natural language understanding?
  - Answering an essay question on an exam?
  - Deciding what to order at a restaurant by reading a menu?
  - Realizing you've been insulted?
  - Appreciating a sonnet?
- As hard as answering "What is artificial intelligence?"

Word senses

- What is the meaning of words?
    - Most words have many different senses:
        - Dog = animal or sausage?



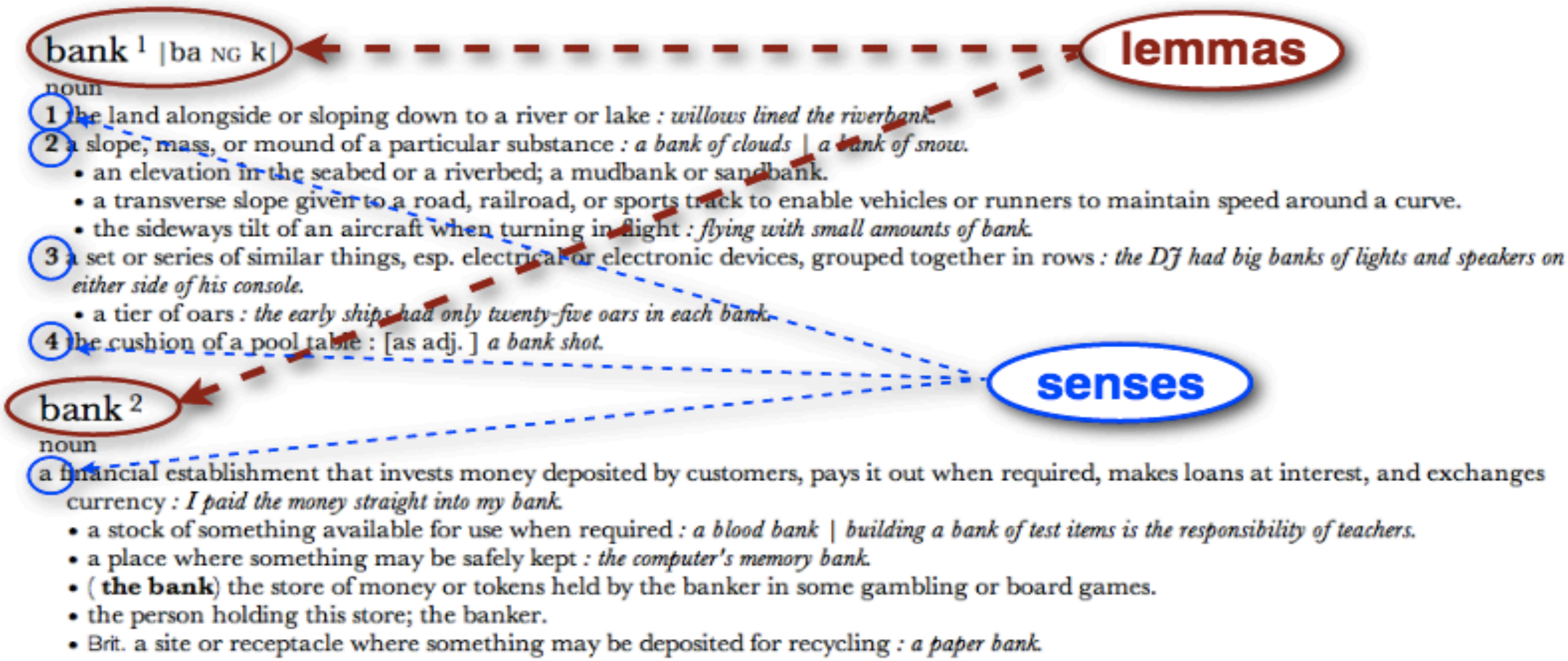


- How are the meanings of different words related?
    - Animal is more general than dog.
    - Money is related to bank.

What does ‘bank mean?’

- -a financial institution
  - (US banks have raised interest rates)
- -a particular branch of a financial institution
  - (the bank on Green Street closes at 5pm)
- -the bank of a river
  - (In 1927, the bank of the Mississippi flooded)
- -a ‘repository’
  - (I donate blood to a blood bank)

Lexicon entries



bank[1] |baNGk|
noun
1 the land alongside or sloping down to a river or lake : *willows lined the riverbank.*
2 a slope, mass, or mound of a particular substance : *a bank of clouds | a bank of snow.*
• an elevation in the seabed or a riverbed; a mudbank or sandbank.
• a transverse slope given to a road, railroad, or sports track to enable vehicles or runners to maintain speed around a curve.
• the sideways tilt of an aircraft when turning in flight : *flying with small amounts of bank.*
3 a set or series of similar things, esp. electrical or electronic devices, grouped together in rows : *the DJ had big banks of lights and speakers on either side of his console.*
• a tier of oars : *the early ships had only twenty-five oars in each bank.*
4 the cushion of a pool table : [as adj.] *a bank shot.*

bank[2]
noun
a financial establishment that invests money deposited by customers, pays it out when required, makes loans at interest, and exchanges currency : *I paid the money straight into my bank.*
• a stock of something available for use when required : *a blood bank | building a bank of test items is the responsibility of teachers.*
• a place where something may be safely kept : *the computer's memory bank.*
• (**the bank**) the store of money or tokens held by the banker in some gambling or board games.
• the person holding this store; the banker.
• Brit. a site or receptacle where something may be deposited for recycling : *a paper bank.*

Some terminology…

- Word forms: runs, ran, running; good, better, best
  - Any, possibly inflected, form of a word
  - (i.e. what we talked about in morphology)
- Lexeme (citation/dictionary form): run
  - A basic word form (e.g. infinitive or singular nominative noun) that is used to represent all forms of the same word. (i.e. the form you'd search for in a dictionary)
- Lemma: RUN (V), GOOD (A), BANK 1 (N), BANK 2 (N)
  - An abstract representation of a word (and all its forms), with a part-of-speech and a set of related word senses. (Often just written (or referred to) as the lemma, perhaps in a different FONT)
- Lexicon:
  - A (finite) list of lexemes

Synonymy

- **Synonymy**: words that have the same meanings or that are closely related in meaning
- E.g. answer/reply – almost/nearly – broad/wide – buy/purchase – freedom/ liberty
- ‘sameness’ is not ‘total sameness’ - only one word would be appropriate in a sentence.
  - E.g. *Sandy only had one answer correct on the test*. (but NOT reply)
- Synonyms differ in formality
  - E.g buy/purchase – automobile/car

Antonymy

- **<u>Antonymy</u>**: words that are opposites in meaning, e.g. hot & cold.

- Types
- *Gradable*= not absolute, question of degree
  - Hot & cold – small & big
- *Non-gradable*:
  - Dead & alive – asleep & awake

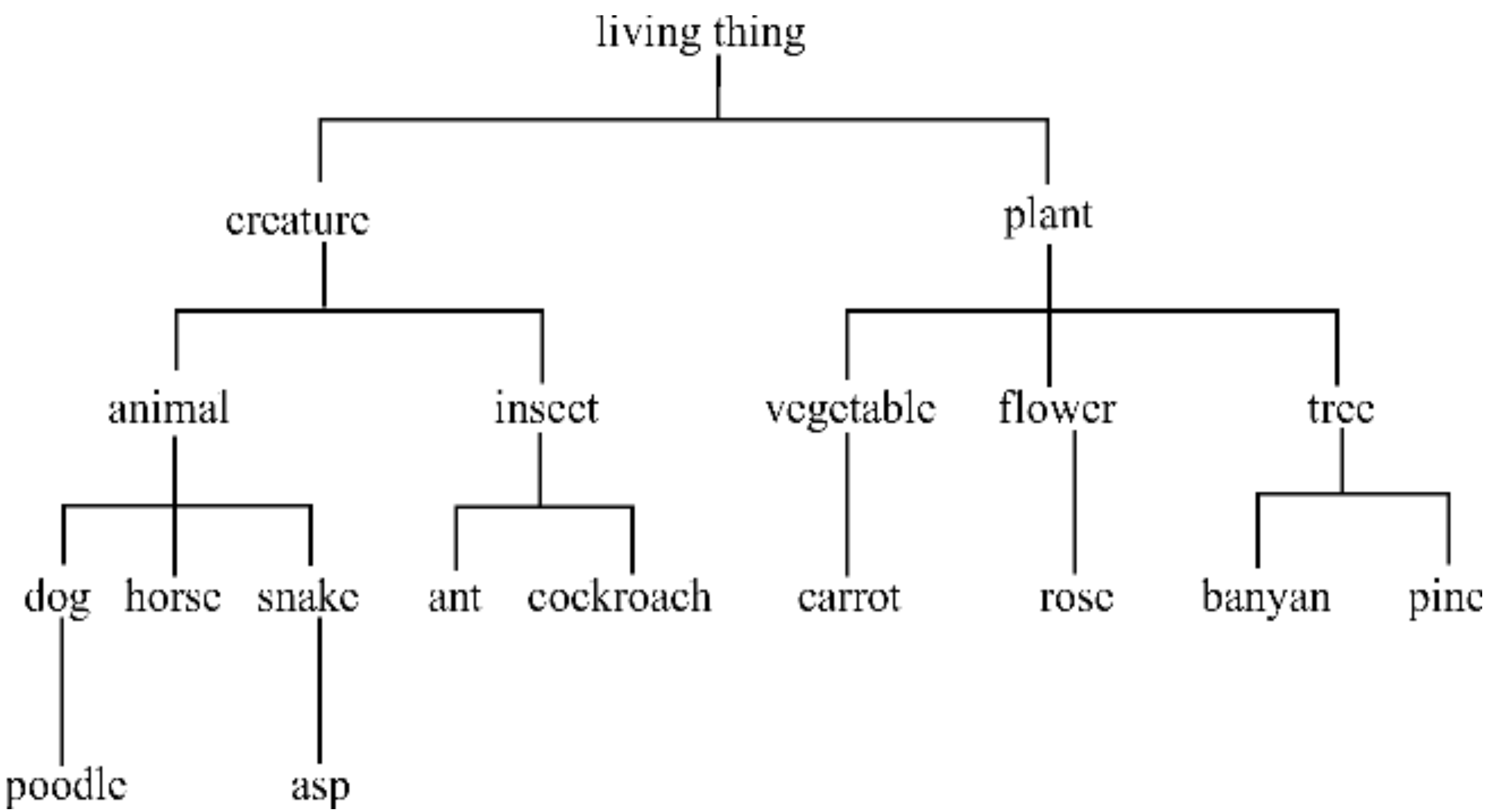
E.g.	happy/sad	married/single
	present/absent	fast/slow

Hyponymy

- **Hyponymy**: Words whose meanings are specific instances of a more general word, i.e. *one thing is included (kind of) in another thing*.
  - e.g. cats and dogs are hyponyms of the word animal.

- Other e.g. daffodil & flower / carrot & vegetable / ant & insect

Hyponymy


living thing
creature
plant
animal
insect
vegetable
flower
tree
dog
horse
snake
ant
cockroach
carrot
rose
banyan
pine
poodle
asp

WordNet

- Very large lexical database of English:
  - 110K nouns, 11K verbs, 22K adjectives, 4.5K adverbs
  - (WordNets for many other languages exist or are under construction)

- Word senses grouped into synonym sets (“synsets”) linked into a conceptual-semantic hierarchy
  - 81K noun synsets, 13K verb synsets, 19K adj. synsets, 3.5K adv synsets
  - Avg. # of senses: 1.23 nouns, 2.16 verbs, 1.41 adj, 1.24 adverbs

- Conceptual-semantic relations: hypernym/hyponym

- Available at http://wordnet.princeton.edu

A WordNet example

WordNet Search - 3.0 - WordNet home page - Glossary - Help

Word to search for: bass | Search WordNet

Display Options: (Select option to change) | Change

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

Noun

- S: (n) **bass** (the lowest part of the musical range)
- S: (n) **bass**, bass part (the lowest part in polyphonic music)
- S: (n) **bass**, basso (an adult male singer with the lowest voice)
- S: (n) sea bass, **bass** (the lean flesh of a saltwater fish of the family Serranidae)
- S: (n) freshwater bass, **bass** (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
- S: (n) **bass**, bass voice, basso (the lowest adult male singing voice)
- S: (n) **bass** (the member with the lowest range of a family of musical instruments)
- S: (n) **bass** (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

Adjective

- S: (adj) **bass**, deep (having or denoting a low vocal or instrumental range) *"a deep voice"; "a bass voice is lower than a baritone voice"; "a bass clarinet"*

Word Similarity

- Synonymy is a binary relation
  - Two words are either synonymous or not
- We want a looser metric
  - Word similarity or
  - Word distance
- Two words are more similar
  - If they share more features of meaning
- Actually these are really relations between **senses**:
  - Instead of saying “bank is like fund”
  - We say
    - Bank1 is similar to fund3
    - Bank2 is similar to slope5
- We’ll compute them over both words and senses

Why Word Similarity

- Spell Checking
- Information retrieval
- Question answering
- Machine translation
- Natural language generation
- Language modeling
- Automatic essay grading

Two Classes of Algorithms

- Thesaurus-based algorithms
  - Based on whether words are "nearby" in WordNet
- Distributional algorithms
  - By comparing words based on their distributional context in corpora

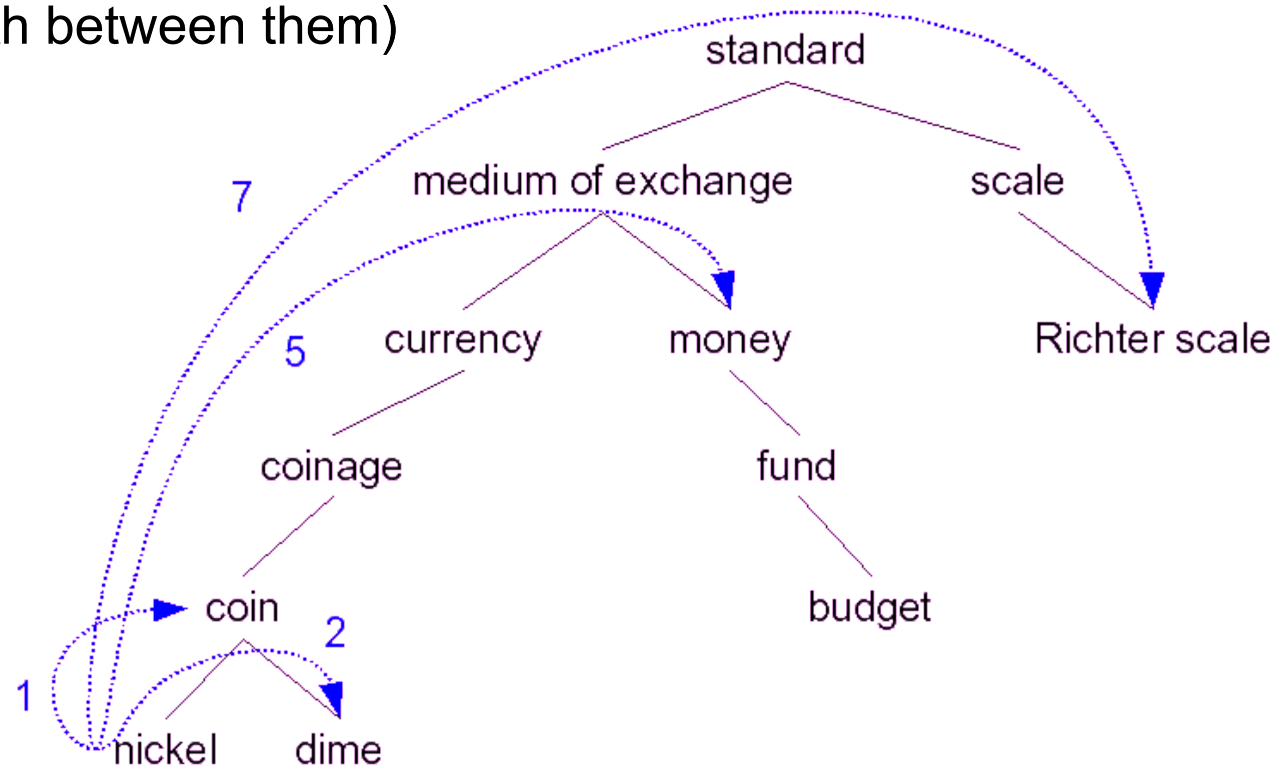
Thesaurus-based Word Similarity

- We could use anything in the thesaurus
  - Example sentences
  - Derivational relations and sentence frames

- Word similarity versus word relatedness
  - Similar words are near-synonyms
  - Related could be related any way
    - Car, gasoline: related, not similar
    - Car, bicycle: similar

Path based similarity

- Two words are similar if nearby in thesaurus hierarchy (i.e. short path between them)

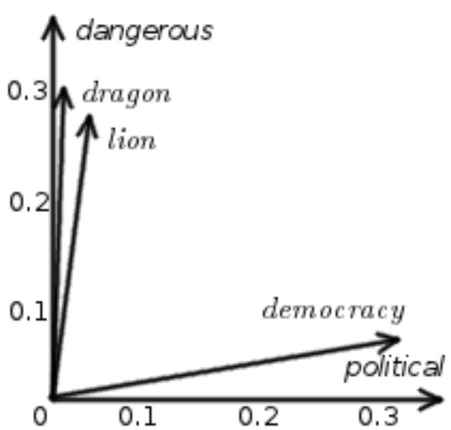


Distributional methods for word similarity

- Firth (1957): "You shall know a word by the company it keeps!"
- An example:
  - A bottle of ***tezgüino*** is on the table
  - Everybody likes ***tezgüino***
  - ***Tezgüino*** makes you drunk
  - We make ***tezgüino*** out of corn.
- Intuition:
  - just from these contexts a human could guess meaning of tezguino
  - So we should look at the surrounding contexts, see what other words have similar context.

Context vector

- Consider a target word *w*
- Suppose we had one binary feature f_i for each of the *N* words in the lexicon v_i
- Which means "word v_i occurs in the neighborhood of *w*"
- w=(f1,f2,f3,…,fN)
- If w=tezguino, v1 = bottle, v2 = drunk, v3 = matrix:
- w = (1,1,0,…)



Intuition

- Define two words by these sparse features vectors
- Apply a vector distance metric
- Say that two words are similar if two vectors are similar

	arts	boil	data	function	large	sugar	summarized	water
apricot	0	1	0	0	1	1	0	1
pineapple	0	1	0	0	1	1	0	1
digital	0	0	1	1	1	0	1	0
information	0	0	1	1	1	0	1	0

Distributional similarity

- So we just need to specify 3 things
  1. How the co-occurrence terms are defined
  2. How terms are weighted
     - (frequency? Logs? Mutual information?)
  3. What vector distance metric should we use?
     - Cosine? Euclidean distance?

Co-occurrence matrices

Word-Word matrix
Sample contexts ± 7 words

sugar, a sliced lemon, a tablespoonful of	**apricot**	preserve or jam, a pinch each of,
their enjoyment. Cautiously she sampled her first	**pineapple**	and another fruit whose taste she likened
well suited to programming on the digital	**computer**.	In finding the optimal R-stage policy from
for the purpose of gathering data and	**information**	necessary for the study authorized in the

	aardvark	computer	data	pinch	result	sugar	…
apricot	0	0	0	1	0	1	
pineapple	0	0	0	1	0	1	
digital	0	2	1	0	1	0	
information	0	1	6	0	4	0	
…	…						

Defining co-occurrence vectors

- We could have windows of neighboring words
  - Bag-of-words
  - We generally remove **stopwords**
- But the vectors are still very sparse
- So instead of using ALL the words in the neighborhood let's just use the words occurring in particular relations

From contexts…

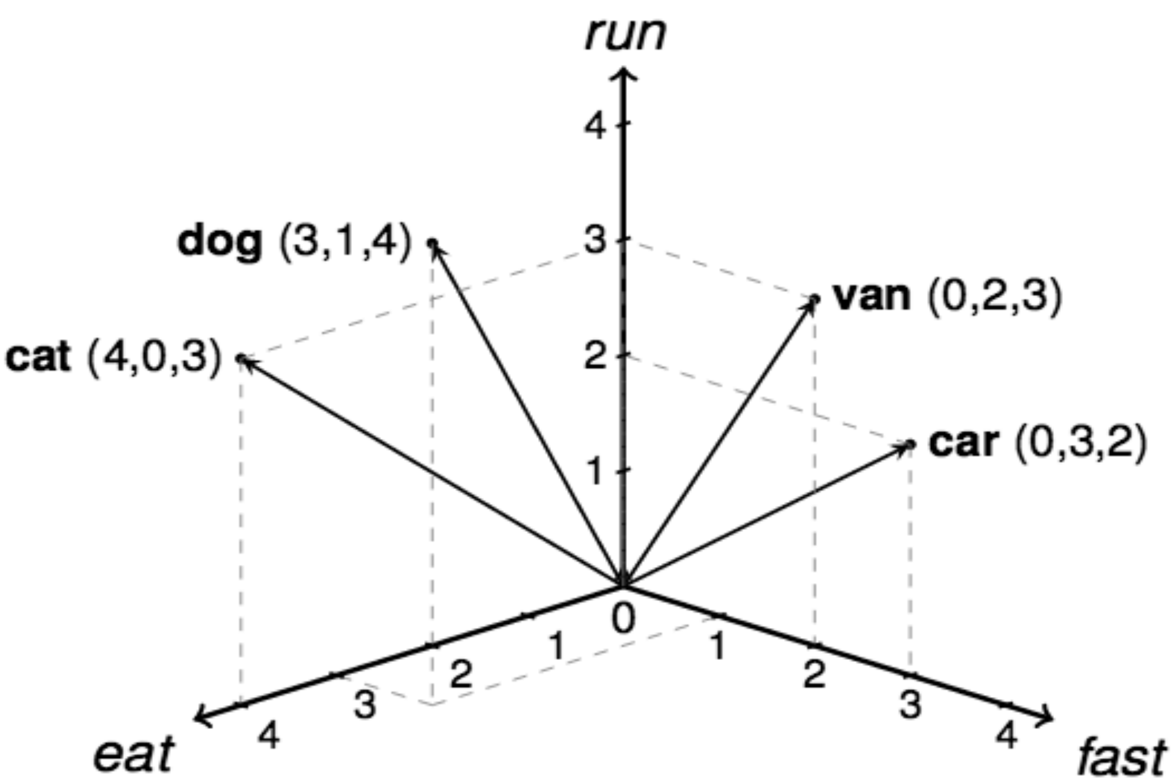
. . . dig a [hole. The	**car**	drove away] leaving behind . . .
. . . to directly [drive the	**car**	wheel angle] 3. Force . . .
. . . celebrity status, [drove fast	**cars**	and partied] with some . . .
. . . but there [are police	**cars**	that chase] you. Each . . .
. . . world of [money, fast	**cars**	and excitement] and, under . . .
. . . to pet [the family's	**cat**	and dog,] who tended . . .
. . . and then [wanted a	**cat**	to eat] the many . . .
. . . murmur is [detectable. The	**cat**	often eats] and drinks . . .
. . . behaviour of [a domestic	**cat**	playing with] a caught . . .
. . . have never [seen a	**cat**	eat so] little and . . .
. . . bank, children [playing with	**dogs**	and a] man leading. . . .
. . . sure you [encourage your	**dog**	to play] appropriate chase . . .
. . . Truth, Lord: [yet the	**dogs**	eat of] the crumbs . . .
. . vegetable material [and enzymes.	**Dogs**	also eat] fruit, berries . . .
. . . hubby once [ate the	**dog**	food and] asked for . . .
. . . were back [at the	**van**	and drove] down to . . .
. . . go down [as the	**van**	drove off.] As he . . .
. . . heavy objects, [driving transit	**vans**	, wiring plugs] and talking . . .
. . . of the [fast food	**van**	being located] outside their . . .
. . . each of [the six	**van**	wheels , and] also under . . .

...to distributional vectors

$$
\begin{array}{cc}
 & \begin{array}{cccccccc} \textit{dog} & \textit{drive} & \textit{eat} & \textit{fast} & \textit{play} & \ldots & \textit{the} & \textit{wheel} \end{array} \\
\begin{array}{l} \textbf{car} \\ \textbf{cat} \\ \textbf{dog} \\ \textbf{van} \end{array} & \begin{pmatrix} 0 & 3 & 0 & 2 & 0 & \vdots & 2 & 1 \\ 1 & 0 & 3 & 0 & 1 & \vdots & 2 & 0 \\ 0 & 0 & 3 & 0 & 2 & \vdots & 2 & 0 \\ 0 & 3 & 0 & 1 & 0 & \vdots & 3 & 1 \end{pmatrix}
\end{array}
$$

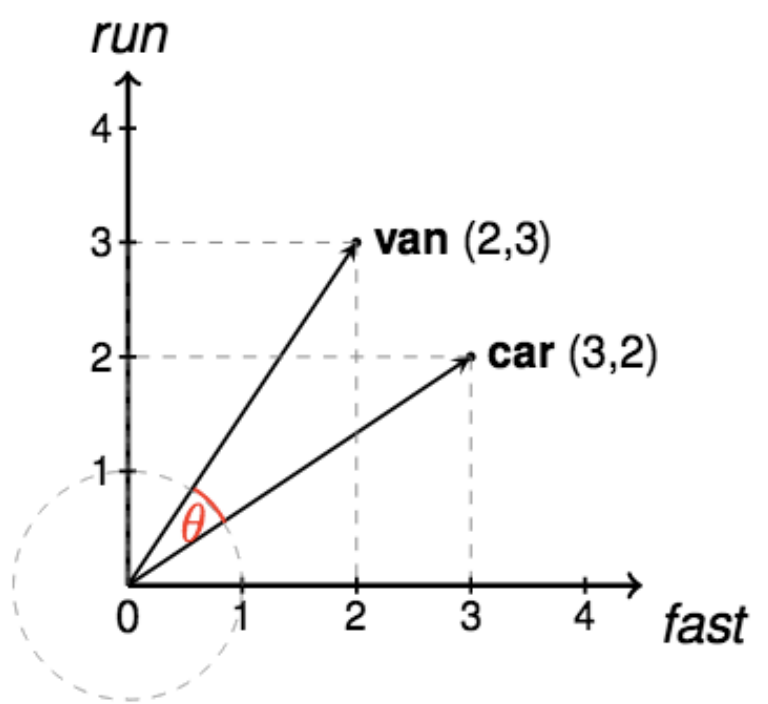
co-occurrence matrix

...to distributional vectors



Measuring vector similarity

$$\text{Cosine} \quad \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}||\mathbf{y}|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$



From vector to semantic similarity

- The Distributional Hypothesis predicts that words with similar distributional vectors are semantically similar

	car	*cat*	*dog*	*van*
car	1			
cat	0.33	1		
dog	0.60	0.94	1	
van	0.92	0.50	0.76	1

cosine similarities

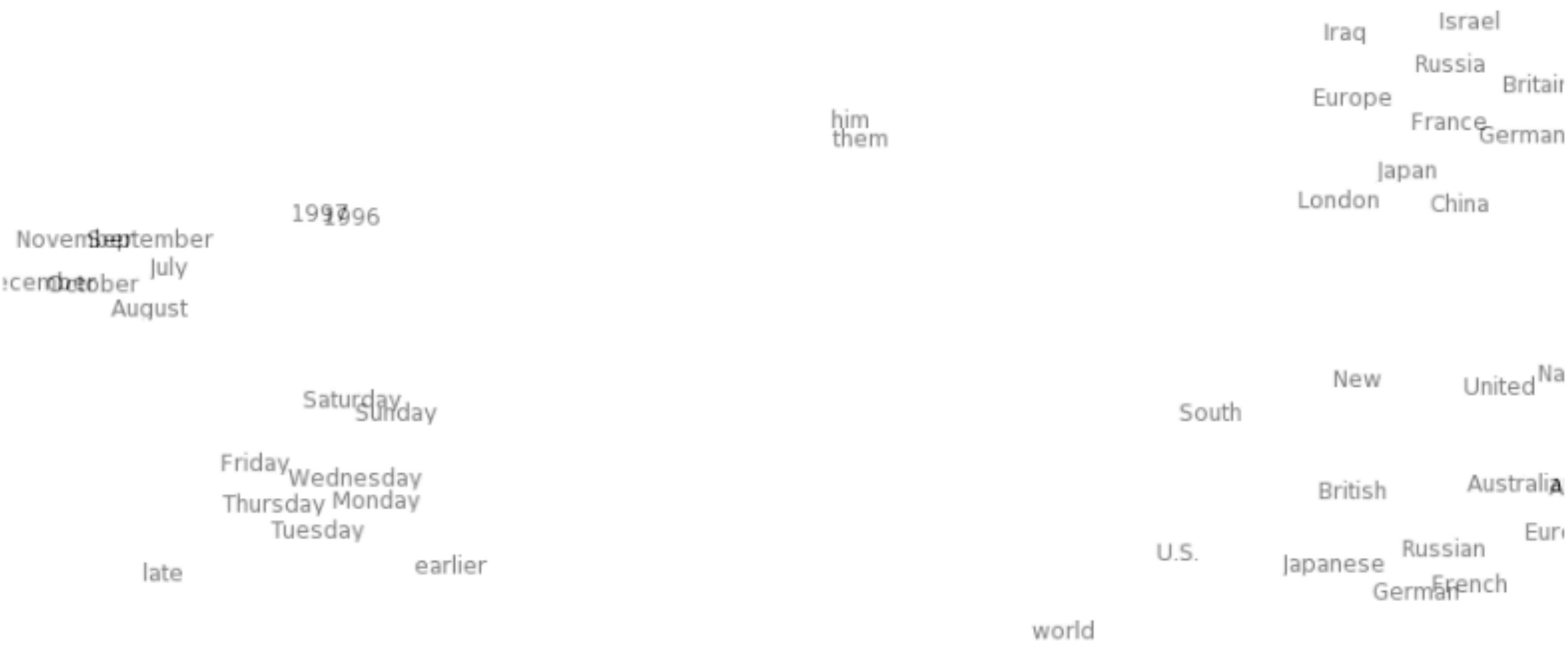
Distributional neighbours from BNC

dog (window size= 2)	
cat	0.77
horse	0.67
fox	0.65
pet	0.63
rabbit	0.61
pig	0.57
animal	0.57
mongrel	0.56
sheep	0.55
pigeon	0.54
deer	0.53
rat	0.53
bird	0.53

good (window size= 2)	
bad	0.68
excellent	0.66
superb	0.48
poor	0.45
improved	0.43
improve	0.43
perfect	0.42
clever	0.42
terrific	0.42
lucky	0.41
smashing	0.41
improving	0.41
wonderful	0.41

Word Vectors

Word Vectors



One application: Question Answering

Why vector models of meaning? computing the similarity between words

"**fast**" is similar to "**rapid**"

"**tall**" is similar to "**height**"

Question answering:

*Q: "How **tall** is Mt. Everest?"*
*Candidate A: "The official **height** of Mount Everest is 29029 feet"*

Another application:
Plagiarism Detection

MAINFRAMES

Mainframes are primarily referred to large computers with rapid, advanced processing capabilities that can execute and perform tasks equivalent to many Personal Computers (PCs) machines networked together. It is characterized with high quantity Random Access Memory (RAM), very large secondary storage devices, and high-speed processors to cater for the needs of the computers under its service.

Consisting of advanced components, mainframes have the capability of running multiple large applications required by many and most enterprises and organizations. This is one of its advantages. Mainframes are also suitable to cater for those applications (programs) or files that are of very high

MAINFRAMES

Mainframes usually are referred those computers with fast, advanced processing capabilities that could perform by itself tasks that may require a lot of Personal Computers (PC) Machines. Usually mainframes would have lots of RAMs, very large secondary storage devices, and very fast processors to cater for the needs of those computers under its service.

Due to the advanced components mainframes have, these computers have the capability of running multiple large applications required by most enterprises, which is one of its advantage. Mainframes are also suitable to cater for those applications or files that are of very large demand

Word Sense Disambiguation (WSD)

Bass: fish

???

Bass: instrument

Word Sense Disambiguation (WSD)

- Given
  - A word in context
  - A fixed inventory of potential word senses
  - Decide which sense of the word this is

- Why? Machine translation, QA, speech synthesis

What does this word mean?

This **plant** needs to be watered each day.
⇒ **living plant**
This **plant** manufactures 1000 widgets each day.
⇒ **factory**

- Word Sense Disambiguation (WSD):
- Identify the sense of content words (noun, verb, adjective)
- in context (assuming a fixed inventory of word senses)

The Data

Sense	Training Examples (Keyword in Context)
?	... company said the *plant* is still operating
?	Although thousands of *plant* and animal species
?	... zonal distribution of *plant* life . ...
?	... to strain microscopic *plant* life from the ...
?	vinyl chloride monomer *plant* , which is ...
?	and Golgi apparatus of *plant* and animal cells
?	... computer disk drive *plant* located in ...
?	... divide life into *plant* and animal kingdom
?	... close-up studies of *plant* life and natural
?	... Nissan car and truck *plant* in Japan is ...
?	... keep a manufacturing *plant* profitable without
?	... molecules found in *plant* and animal tissue
?	... union responses to *plant* closures . ...
?	... animal rather than *plant* tissues can be
?	... many dangers to *plant* and animal life
?	company manufacturing *plant* is in Orlando ...
?	... growth of aquatic *plant* life in water ...
?	automated manufacturing *plant* in Fremont ,
?	... Animal and *plant* life are delicately
?	discovered at a St. Louis *plant* manufacturing
?	computer manufacturing *plant* and adjacent ...
?	... the proliferation of *plant* and animal life
?	... ...

WSD as a learning problem

- Supervised:
  - You have a (large) corpus annotated with word senses
  - Here, WSD is a standard supervised learning task

- Semi-supervised approaches:
  - You only have very little annotated data (and a lot of raw text)
  - Here, WSD is a semi-supervised learning task

- Unsupervised approaches:
  - You don’t have any annotated data.

Implementing a WSD Classifier

- Basic insight: The sense of a word in a context depends on the words in its context

- Features:
- - Which words in context: all words, all/some content words
- - How large is the context? sentence, prev/following 5 words
- - Do we represent context as bag of words (unordered set of words) or do we care about the position of words (preceding/following word)?
- - Do we care about POS tags?
- - Do we represent words as they occur in the text or as their lemma (dictionary form)?

Supervised Machine Learning Approaches

- Supervised machine learning approach:
  - a training corpus of words tagged in context with their sense
  - used to train a classifier that can tag words in new text
- Summary of what we need:
  - the **tag set** (“sense inventory”)
  - the **training corpus**
  - A set of **features** extracted from the training corpus
  - A **classifier**

Supervised WSD 1: WSD Tags

- What's a tag?
  A dictionary sense?

- For example, for WordNet an instance of "bass" in a text has 8 possible tags or labels (bass1 through bass8).

8 Senses of "bass" in WordNet

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

Supervised WSD 2: Get a corpus

- Lexical sample task:
  - *Line-hard-serve* corpus - 4000 examples of each
  - *Interest* corpus - 2369 sense-tagged examples
- All words:
  - **Semantic concordance**: a corpus in which each open-class word is labeled with a sense from a specific dictionary/thesaurus.
    - SemCor: 234,000 words from Brown Corpus, manually tagged with WordNet senses
    - SENSEVAL-3 competition corpora - 2081 tagged word tokens

Feature vectors

- A simple representation for each observation
  (each instance of a target word)
  - **Vectors** of sets of feature/value pairs
  - Represented as an ordered list of values
  - These vectors represent, e.g., the window of words around the target

Two kinds of features in the vectors

- **Collocational** features and **bag-of-words** features
  - **Collocational**
    - Features about words at **specific** positions near target word
      - Often limited to just word identity and POS
  - **Bag-of-words**
    - Features about words that occur anywhere in the window (regardless of position)
      - Typically limited to frequency counts

Examples

- Example text (WSJ):

  An electric guitar and **bass** player stand off to one side not really part of the scene

- Assume a window of +/- 2 from the target

Examples

- Example text (WSJ)

  An electric guitar and **bass** player stand off to one side not really part of the scene,

- Assume a window of +/- 2 from the target

Collocational Features

- Position-specific information about the words and collocations in window
- $\boxed{\text{guitar}}\boxed{\text{and}}$ bass $\boxed{\text{player}}\boxed{\text{stand}}$

$$\left[w_{i-2}, \text{POS}_{i-2}, w_{i-1}, \text{POS}_{i-1}, w_{i+1}, \text{POS}_{i+1}, w_{i+2}, \text{POS}_{i+2}, w_{i-2}^{i-1}, w_{i}^{i+1}\right]$$

[guitar, NN, and, CC, player, NN, stand, VB, and guitar, player stand]

- word 1,2,3 grams in window of ±3 is common

Bag-of-Words Features

- “an unordered set of words” – position ignored
- Counts of words occur within the window.
- First choose a vocabulary
- Then count how often each of those terms occurs in a given window
  - sometimes just a binary “indicator” 1 or 0

Co-Occurrence Example

- Assume we've settled on a possible vocabulary of 12 words in "bass" sentences:

[*fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band*]

- The vector for:

  guitar and bass player stand

  [0,0,0,1,0,0,0,0,0,0,1,0]

Classification: definition

- *Input*:
  - a word w and some features f
  - a fixed set of classes $C = \{c_1, c_2, \ldots, c_J\}$
- *Output*: a predicted class $c \in C$

Classification Methods: Supervised Machine Learning

- *Input:*
  - a word *w in a text window d (which we'll call a "document")*
  - a fixed set of classes $C = \{c_1, c_2,\ldots, c_J\}$
  - A training set of m hand-labeled text windows again called "documents" $(d_1,c_1),\ldots,(d_m,c_m)$
- *Output:*
  - a learned classifier $\gamma:d \rightarrow c$

Classification Methods: Supervised Machine Learning

- Any kind of classifier
  - **Naive Bayes**
  - Logistic regression
  - Neural Networks
  - Support-vector machines
  - k-Nearest Neighbors
  - …

Applying Naive Bayes to WSD

- P(c) is the prior probability of that sense
  - Counting in a labeled training set.
- P(w|c) conditional probability of a word given a particular sense
  - P(w|c) = count(w,c)/count(c)
- We get both of these from a tagged corpus like SemCor, Senseval, etc.

- Can also generalize to look at other features besides words.
  - Then it would be P(f|c)
    - Conditional probability of a feature given a sense

We assume that each feature is conditionally independent of all the other features:

$$
\begin{aligned}
\hat{s} = \quad & \arg\max_{s \in S} P(s|\vec{f}) & \\
= \quad & \arg\max_{s \in S} \frac{P(\vec{f}|s)P(s)}{P(\vec{f})} & \text{Bayes} \\
= \quad & \arg\max_{s \in S} P(\vec{f}|s)P(s) & P(\vec{f}) \text{ is fixed} \\
\approx \quad & \arg\max_{s \in S} P(s) \prod_{j=1}^{n} P(f_j|s) & \text{cond. independence}
\end{aligned}
$$

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(w \mid c) = \frac{count(w,c)+1}{count(c)+|V|}$$

	Doc	Words	Class
Training	1	fish smoked fish	f
	2	fish line	f
	3	fish haul smoked	f
	4	guitar jazz line	g
Test	5	line guitar jazz jazz	?

V = {fish, smoked, line, haul, guitar, jazz}

Priors:

$P(f) = \frac{3}{4}$

$P(g) = \frac{1}{4}$

Choosing a class:

P(f|d5) $\propto$ 3/4 * 2/14 * $(1/14)^2$ * 1/14 ≈ 0.00003

Conditional Probabilities:

P(line|*f*) = (1+1) / (8+6) = 2/14

P(guitar|*f*) = (0+1) / (8+6) = 1/14

P(jazz|*f*) = (0+1) / (8+6) = 1/14

P(line|*g*) = (1+1) / (3+6) = 2/9

P(guitar|*g*) = (1+1) / (3+6) = 2/9

P(jazz|*g*) = (1+1) / (3+6) = 2/9

P(g|d5) $\propto$ 1/4 * 2/9 * $(2/9)^2$ * 2/9 ≈ 0.0006

References

- Dan Jurafsky, WordNet, word similarity, sense relations, 2007
- Paula Matuszek, Mary-Angela Papalaskari, Semantics and semantic analysis, Spring, 2005
- Christopher Manning, Dan Klein, slides
- Dan Jurafsky, Word Sense Disambiguation slides
- Alex Lascarides, Lexical Semantics slides